9 788883 032950

Tools for Identifying Biodiversity:
Progress and Problems

Pier Luigi Nimis and Régine Vignes Lebbe (eds.)

2010

Digitized by the Internet Archive
in 2026

https://archive.org/details/toolsforidentify0000unse

Tools for Identifying Biodiversity:
Progress and Problems

Pier Luigi Nimis and Régine Vignes Lebbe (eds.)

EUT

9 788883 032950

Tools for Identifying Biodiversity:
Progress and Problems

Pier Luigi Nimis and Régine Vignes Lebbe (eds.)

Proceedings of the International Congress
Paris, September 20-22, 2010
Muséum national d'Histoire naturelle – Grand Amphithéâtre

© All content is copyrighted by the individual authors of the contributions and licensed under the Creative Commons Attribution-Share-Alike License (CC by-sa 3.0).
EUT 2010.
Information in this book does not represent the opinion of the European Community and the European Community is not responsible for any use that might be made of it.

ISBN 978-88-8303-295-0

EUT - Edizioni Università di Trieste
Via E. Weiss, 21 – 34128 Trieste
http://eut.units.it

Cover and layout: Rodolfo Riccamboni, Divulgando S.r.l. – www.divulgando.eu
Printed in Italy by Studio Pixart S.r.l., Via I Maggio, 8, I – 30020 Quarto d'Altino, VENEZIA.

Scientific Committee

Nicolas Bailly (Philippines)
Daniel Barthélémy (France)
Frank Bisby (United Kingdom)
Walter Berendsohn (Germany)
Gudmundur Gudmundsson (Iceland)
Gregor Hagedorn (Germany)
Bill Hominick (United Kingdom)
Christian Kittl (Austria)
Maurizio Casiraghi (Italy)
Noel Conruyt (La Réunion, France)
Martin Grube (Austria)
Massimo Labra (Italy)
Karol Marhold (Slovak Republic)
Anastasia Miliou (Greece)
Andrea M. Mulrenin (Austria)
Pier Luigi Nimis (Italy)
Bob Press (United Kingdom)
Dave Roberts (United Kingdom)
Peter Schalk (Netherlands)
Edwin van Spronsen (Netherlands)
Régine Vignes Lebbe (France)

Organising Committee

Coordinators:
Pier Luigi Nimis
Régine Vignes Lebbe

Members:
Léa Bled
Florian Causse
Vanessa Demanoff
Zoulika Labghiel
Visotheary Rivière-Ung
Stefano Martellos
Rodolfo Riccamboni
Maxime Venin

Foreword

The correct identification of organisms is fundamental not only for the assessment and the conservation of biodiversity, but also in agriculture, forestry, the food and pharmaceutical industries, forensic biology, and in the broad field of formal and informal education at all levels. However, since the first Meeting of the Systematics Association on «Biological identification with computers», in 1973, few scientific events have been dedicated to this subject. Furthermore, taxonomists, workers in biodiversity informatics, and the large community of users are rarely all gathered together.

Since the 1990s, the number of projects developing information repositories has greatly increased: Fishbase, GBIF, Species 2000, OBIS, EuroMed-PlantBase, Fauna Europaea, EoL etc. to cite only some of them. Until now, identification tools were poorly represented in such systems. This is already changing, and Fishbase is a good example illustrating the need to include identification facilities with biodiversity databases, and to adapt the keys to different types of users. International conferences on biodiversity research, tools and methods using ICT, are becoming more and more numerous. In the last decades, important advances have taken place in the ways identification is carried out, from molecular and biochemical methods of rapid identification to the development of interactive identification systems based on morpho-anatomical data. The effort to propose and to popularize identification tools using all types of biological characters (sequences, morphology, images, sounds etc.) must be continued.

The event «Tools for identifying biodiversity: progress and problems» offers an opportunity to provide an overview of recent advances in this field. It aims at stimulating integration of existing methods and systems, fostering communication amongst different research groups, and laying the foundations for integrated projects in the next decade. The congress was organised jointly by three large European projects dedicated to biodiversity and/or biological identification: KeyToNature, EDIT (European Distributed Institute of Taxonomy), and STERNA (Semantic Web-based Thematic European Reference Network Application).

The scientific program of the congress was subdivided into four sessions:

- *Interactive identification tools based on morpho-anatomical data*
- *Molecular and biochemical methods for the identification of organisms*
- *Identification and education*
- *Industrial and practical applications of the new identification tools: case-studies and markets*

In this book, the reader will find short presentations of current and upcoming projects (EDIT, KeyToNature, STERNA, Species 2000, Fishbase, BHL, ViBRANT, etc.), plus a large panel of short articles on software, taxonomic applications, use of e-keys in the educational field, and practical applications. Single-access keys are now available on most recent electronic devices; the collaborative and semantic web opens new ways to develop and to share applications; the automatic processing of molecular data and images is now based on validated systems; identification tools appear as an efficient support for environmental education and training; the monitoring of invasive and protected species and the study of climate change require intensive identifications of specimens, which opens new markets for identification research.

Pier Luigi Nimis, Régine Vignes Lebbe

Trieste – Paris, September 2010

Table of Contents

Devising the EDIT Platform for Cybertaxonomy 1
Walter G. Berendsohn

Descriptive Data in the EDIT Platform for Cybertaxonomy 7
Maxime Venin, Agnes Kirchhoff, Hélène Fradin, Anton Güntsch, Niels Hoffmann, Andreas Kohlbecker, Elise Kuntzelmann, Ôna Maiocco, Andreas Müller, Régine Vignes Lebbe, Walter G. Berendsohn

An online authoring and publishing platform for field guides and identification tools 13
Gregor Hagedorn, Gisela Weber, Andreas Plank, Mircea Giurgiu, Andrei Homodi, Cornelia Veja, Gerd Schmidt, Pencho Mihnev, Manol Roujinov, Dagmar Triebel, Robert A. Morris, Bernhard Zelazny, Edwin van Spronsen, Peter Schalk, Christian Kittl, Robert Brandner, Stefano Martellos, Pier Luigi Nimis

A search tool for the digital biodiversity resources of *KeyToNature* 19
Mircea Giurgiu, Andrei Homodi, Cornelia Veja, Gregor Hagedorn, Pier Luigi Nimis

Developing Web-based Search Portals on Birds for Different Target Groups 25
Renate Steinmann, Andreas Strasser, Andrea Mulrenin, Amy Trayler, Sander Pieterse, Ivan Teage, Michael De Giovanni, John J. Borg, Noel Zammit

Simple Identification Tools in FishBase 31
Nicolas Bailly, Rodolfo Reyes Jr., Rachel Atanacio, Rainer Froese

The Catalogue of Life: towards an integrative taxonomic backbone for biodiversity 37
Frank A. Bisby, Yuri R. Roskov

BHL-EUROPE: Biodiversity Heritage Library for Europe 43
Jana Hoffmann, Henning Scholz

A Pan-European Species-directories Infrastructure (PESI) 49
Yde de Jong

ViBRANT–Virtual Biodiversity Research and Access Network for Taxonomy 53
Dave Roberts, Vince Smith

Identifications in BioPortals™ 55
Wouter Addink, Edwin van Spronsen, Peter H. Schalk

Types of identification keys 59
Gregor Hagedorn, Gerhard Rambold, Stefano Martellos

Learning, Identifying, Sharing 65
Philippe A. Martin, Noël Conruyt, David Grosser

Identification with iterative nearest neighbors using domain knowledge 71
David Grosser, Noël Conruyt, Henri Ralambondrainy

A MediaWiki implementation of single-access keys 77
Gregor Hagedorn, Bob Press, Sonia Hetzner, Andreas Plank, Gisela Weber, Sabine von Mering, Stefano Martellos, Pier Luigi Nimis

Simple matrix keys from Excel spreadsheets 83
Gregor Hagedorn, Mircea Giurgiu, Andrei Homodi

Wiki keys on mobile devices 89
Gisela Weber, Gregor Hagedorn

A Wiki-based Key to Garden and Village Birds 95
Tomi Trilar

Wiki-keys for the ferns of the Flora of Equatorial Guinea 99
Francisco Cabezas, Carlos Aedo, Patricia Barberá, Manuel De la Estrella, Maximiliano Fero, Mauricio Velayos

MyKey: a server-side software to create customized decision trees 107
David Gérard, Régine Vignes Lebbe

Xper²: managing descriptive data from their collection to e-monographs 113
Visotheary Ung, Florian Causse, Régine Vignes Lebbe

FRIDA 3.0 Multi-authored digital identification keys in the Web 115
Stefano Martellos

Flora Bellissima, an expert software to discover botany and identify plants 121
Thierry Pernot, Daniel Mathieu

Modifiable digital identification keys 127
Edwin van Spronsen, Stefano Martellos, Dennis Seijts, Peter Schalk, Pier Luigi Nimis

The Open Key Player: A new approach for online interaction and user-tracking in identification keys 133
Mircea Giurgiu, Andrei Homodi, Edwin van Spronsen, Stefano Martellos, Pier Luigi Nimis

Improvement of identification keys by user-tracking 137
Gerd Schmidt, Mircea Giurgiu, Sónia Hetzner, Fred Neumann

ARIES: an expert system supporting legislative tasks. Identifying animal materials using the Linnaeus II software 145
Leo W.D. van Raamsdonk

An integrated system for producing user-specific keys on demand: an application to Italian lichens 151
Juri Nascimbene, Stefano Martellos, Pier Luigi Nimis

"Flora Italiana Digitale": an interactive identification tool for the Flora of Italy 157
Riccardo Guarino, Sabina Addamiano, Marco La Rosa, Sandro Pignatti

eFlora and DialGraph, tools for enhancing identification processes in plants 163
Fernando Sánchez Laulhé, Cecilio Cano Calonge, Antonio Jiménez Montaño

A catalogue of bird bones: an exercise in semantic web practice 171
Gudmundur Gudmundsson, Seth D. Brewington, Thomas H. McGovern, Aevar Petersen

Anthos.es: 10 years showing Spanish plant diversity information in the Internet..... 177
Leopoldo Medina, Carlos Aedo

An interactive tool for the identification of airborne and food fungi..... 183
Giovanna Cristina Varese, Antonella Anastasi, Samuele Voyron, Valeria Filipello Marchisio

The Estonian eFlora 189
Tiina Randlane, Malle Leht, Andres Saag

Keys to plants and lichens on smartphones: Estonian examples 195
Andres Saag, Tiina Randlane, Malle Leht

IIKC: An Interactive Identification Key for female *Culicoides* (Diptera: Ceratopogonidae) from the West Palearctic region..... 201
Bruno Mathieu, Catherine Cêtre-Sossah, Claire Garros, David Chavernac, Thomas Balenghien, Régine Vignes Lebbe, Visotheary Ung, Ermanno Candolfi, Jean-Claude Delécolle

Indochinese bamboos: biodiversity informatics to assist the identification of "vernacular taxa"..... 207
My Hanh Diep Thi, Régine Vignes Lebbe, Ha Phuong Nguyen, Bich Loan Nguyen Thi

Identification tools as part of Feedsafety research: the case of ragwort 213
Leo W.D. van Raamsdonk, Patrick Mulder, Michel Uiterwijk

Two identification tools applied on Mascarene's corals genera (Xper^2) and species (IKBS)..... 217
Yannick Geynet, Noël Conruyt, David Grosser, Gérard Faure, David Caron

Interactive, illustrated, plant identification keys: an example for the Portuguese flora 219
Maria Helena Abreu Silva, Rosa Maria Ferreira Pinho, Lísia Graciete, Martins Pereira Lopes, Paulo Cardoso da Silveira

The ORCHIS software used to identify 100 orchids species of Lao PDR 221
Pierre Bonnet, André Schuiteman, Boukhaykhone Svengsuksa, Daniel Barthélémy, Vichith Lamxay, Soulivanh Lanorsavanh, Khamfa Chanthavongsa, Pierre Grard

A collaborative and distributed identification tool for plants 223
Philippe Laroche

Alternative 2D and 3D Form Characterization Approaches to the Automated Identification of Biological Species..... 225
Norman MacLeod

VeSTIS: A Versatile Semi-Automatic Taxon Identification System from Digital Images 231
Nikos Nikolaou, Pantelis Sampaziotis, Marilena Aplikioti, Andreas Drakos, Ioannis Kirmitzoglou, Marina Argyrou, Nikos Papamarkos, Vasilis J. Promponas

Iterative Search with Local Visual Features for Computer Assisted Plant Identification 237
Wajih Ouertani, Pierre Bonnet, Michel Crucianu, Nozha Boujemaa, Daniel Barthélémy

Image data banks and geometric morphometrics 243
Anna Loy, Dennis E. Slice

Outline analysis for identifying *Limodorum* species from seeds 249
Sara Magrini, Sergio Buono, Emanuele Gransinigh, Massimiliano Rempicci, Silvano Onofri, Anna Scoppola

Geometric morphometrics as a tool to resolve taxonomic problems: the case of *Ophioglossum* species (ferns) 251
Sara Magrini, Anna Scoppola

Geometric morphometric analysis as a tool to explore covariation between shape and other quantitative leaf traits in European white oaks ... 257
Vincenzo Viscosi, Anna Loy, Paola Fortini

Landmark based morphometric variation in Common dolphin (*Delphinus delphis L.*,1758) 263
Paola Nicolosi, Anna Loy

DNA barcoding: theoretical aspects and practical applications 269
Maurizio Casiraghi, Massimo Labra, Emanuele Ferri, Andrea Galimberti, Fabrizio De Mattia

Strength and Limitations of DNA Barcode under the Multidimensional Species Perspective 275
Valerio Sbordoni

DNA Barcoding and Phylogeny of Patellids from Asturias (Northern Spain) 281
Yaisel Juan Borrell, Fernando Romano, Emilia Vázquez, Gloria Blanco, Jose Antonio Sánchez Prado

Molecular Identification of Italian Mouse-eared Bats (genus *Myotis*) 289
Andrea Galimberti, Adriano Martinoli, Danilo Russo, Mauro Mucedda, Maurizio Casiraghi

Identifying algal symbionts in lichen symbioses 295
Martin Grube, Lucia Muggia

Identification of polymorphic species within groups of morphologically conservative taxa: combining morphological and molecular techniques 301
Kim Larsen, Elsa Froufe

Coffee species and varietal identification 307
Patrizia Tornincasa, Michela Furlan, Alberto Pallavicini, Giorgio Graziosi

Mislabelling in megrims: implications for conservation 315
Victor Crego-Prieto, Daniel Campo, Juliana Perez, Eva Garcia-Vazquez

Seeds in subtribe Orchidinae (Orchidaceae):
the best morphological tool to support molecular analyses 323
Roberto Gamarra, Emma Ortúñez, Ernesto Sanz, Iris Esparza, Pablo Galán

Lentils biodiversity: the characterization of two local landraces 327
Vincenzo Viscosi, Manuela Ialicicco, Mariapina Rocco, Dalila Trupiano, Simona Arena, Donato Chiatante, Andrea Scaloni, Gabriella Stefania Scippa

A model study for tardigrade identification 333
Roberto Bertolani, Lorena Rebecchi, Michele Cesari

DNA Barcoding of Philippine plants 341
Esperanza Maribel G. Agoo

Molecular and ecophysiological characterisation of the Tunisian bee:
Apis mellifera intermissa 343
Mohamed Chouchene, Naima Barbouche, Lionel Garnery, Michel Baylac

Biological identifications through mitochondrial and nuclear molecular markers: the case of commercially important crabs from Indian EEZ 345
Sherine Sonia Cubelio, K. K. Bineesh, K. Raj, Suraj Tewari, Achamveettil Gopalakrishnan, Valaparambil Saidumohammad Basheer, Wazir Singh Lakra

Barcoding Fauna Bavarica – Capturing Central European
Animal Diversity 347
Lars Hendrich, Michael Balke, Gerhard Haszprunar, Axel Hausmann, Paul Hebert, Stefan Schmidt

Molecular techniques for identifying North Sea fauna 349
Thomas Knebelsberger, Sandra Ditzler, Silke Laakmann, Inga Mohrbeck, Michael J. Raupach

DNA Bank Network – connecting biological collections and sequence databases by long-term DNA storage with online accession 351
Matthias Geiger, Nicolas Straube

Mitochondrial DNA sequences for forensic identification of the endangered whale shark, *Rhincodon typus* (Smith, 1828): A Case study 353
Kavungal Abdulkhadar Sajeela, Chandran Rakhee, Janardanan Nair Rekha, Achamveettil Gopalakrishnan, Valaparambil Saidumohammad Basheer, Joe Kizhakkudan Shoba, Kizhakkudan Joe, Wazir Singh Lakra

An assignment-based e-learning course on the use
of *KeyToNature* e-keys 355
Pencho Mihnev, Nadezhda Raycheva

User needs for interactive identification tools to organisms
employed in the EU-Project *KeyToNature* 361
Astrid Tarkus, Emanuel Maxl, Christian Kittl

Teaching biodiversity with online identification tools from *KeyToNature*:
a comparative study 367
Felicia Boar, Adelhaida Kerekes

Digital Tools in the Botanical Garden of Madrid 373
Marina Ferrer, Esther García

Use of *KeyToNature* Identification Tools in the Schools of Slovenia 379
Irena Kodele Krašna

New key-tools for pollen identification in research and education 383
Jade Dupont, Nathalie Combourieu Nebout, Jean-Pierre Cazet, Florian Causse, Régine Vignes Lebbe

The UK urban tree survey 389
Bob Press

Tree School – A new Innovation for Science and Education 395
Della Hopkins, Karen James

Engaging Schools in Cutting Edge Science: From the Educator's Perspective 401
Adrian Richardson, Della Hopkins

Educational or emotional languages? An interactive experiment with the Lucanian flora (S-Italy) 405
Riccardo Guarino, Patrizia Menegoni, Sandro Pignatti

Online sharing educational content on biodiversity topics: a case study from organic agriculture and agroecology 411
Nikos Palavitsinis, Nikos Manouselis, Kostas Kastrantas, John Stoitsis, Xenofon Tsilibaris

JSTOR Plant Science 417
Michael Sean Gallagher

ecoBalade: Towards a workflow for Citizen Science Nature Trails 419
Julie Chabalier, Khaled Talbi, Patrick Peters, Amandine Sahl, Olivier Coullet, Olivier Assunçao, Olivier Rovellotti

Electronic data recording tools and identifying species in the field 421
Alexander Kroupa, Anke Hoffmann, Juan Carlos Monje, Christoph L. Häuser

Cost Assessment of the Field Measurement of Biodiversity: a Farm-scale Case Study 423
Stefano Targetti, Davide Viaggi, David Cuming

Markets for biodiversity information products: real or imaginary? 429
Bill Hominick, Peter Schalk

A Basic Business Model for Commercial Application of Identification Tools 437
Christian Kittl, Peter Schalk, Nicola Dorigo Salamon, Stefano Martellos

Keys to Nature: A test on the iPhone market 445
Rodolfo Riccamboni, Alessio Mereu, Chiara Boscarol

Author Index 451

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 1-6.
ISBN 978-88-8303-295-0. EUT, 2010.

Devising the EDIT Platform for Cybertaxonomy

Walter G. Berendsohn

Abstract — This contribution describes the original ideas and preparatory work that led to the implementation of the EDIT Platform for Cybertaxonomy, a computing environment supporting the entire taxonomic workflow. It also briefly describes the current state of development of the project, which will end its EU-funded period in February, 2011.

Index Terms — biodiversity informatics, cybertaxonomy, taxonomic computing, taxonomy.

1 Introduction

Taxonomic research is traditionally a highly collaborative endeavour. The EU project EDIT (European Distributed Institute of Taxonomy) brings together a consortium including most of the largest European natural history museums. EDIT aims at integrating taxonomic research at multiple levels: research policies, collection management, training, outreach and public relations, and research infrastructure.

Natural history museums are information and knowledge institutions. In EDIT, the relatively new area of information technologies was seen as a major chance to integrate the activities of the project partners. Therefore, a third of the project funding was dedicated to create the “Internet Platform for Cybertaxonomy”.

W. G. Berendsohn is with the Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin.
Instead of filling this page with co-authors, the author here lists the collaborators in the EDIT work-package 5 “Internet Platform for Cybertaxonomy” in a comprehensive form:
The task leaders in the work package were Wieslaw Bogdanowicz, Museum of Invertebrate Zoology, Polish Academy of Sciences (MIZPAN); Andras Gubanyi, Hungarian Natural History Museum, Budapest (HNHM); Anton Güntsch, BGBM; Christoph Häuser, State Museum for Natural History, Stuttgart (SMNS) (now at MfN); Mark Jackson, Royal Botanic Gardens, Kew (RBGK); Jorge Lobo, Museo Nacional de Ciencias Naturales, Madrid (CSIC); Karol Marhold, Institute of Botany, Slovak Academy of Sciences (IBSAS); Patricia Mergen, Royal Museum for Central Afrika, Tervuren (RMCA); Martin Pullan, Royal Botanic Garden, Edinburgh (RBGE); Henning Scholz, Museum für Naturkunde, Berlin (MfN); Jane Smith, Natural History Museum, London (NHML); Eduard Stloukal (CUB) and Régine Vignes, Université Pierre et Marie Curie Paris 6 (UPMC). The EDIT development team was first led by Markus Döring and from the second year on jointly by Andreas Kohlbecker and Andreas Müller (BGBM Berlin). Team members were (alphabetically; independent of the time span they worked for EDIT): Anahit Badadshanjan (BGBM), Elek Bozóky-Szeszich (HNHM), Garin Cael (RMCA),Pepe Ciardelli (BGBM), Ben Clark (RBGK), Nils Clark-Bernhard (independent), James Davy (RMCA), Marco Figuidero (NHML), Helene Fradin (UPMC), Giovanni

Thirteen institutions from 8 countries directly participated in the workpackage elaborating the Platform, 7 institutions were involved in software development (programming), with a total of 25 developers (12 concurrent) busy forging the code.

2 Arriving at the Specification

At the starting point of the project in March 2005 we knew that we had a well-resourced project, but we also knew that software development may exceed all cost expectations, especially when carried out in large cooperative projects and (worse) when relatively new technologies were to be used. We were determined not to re-invent the wheel and reuse existing software as much as possible.

The first project year was spent analysing existing software, standards, infrastructures at partner institutions, and requirements from taxonomists, especially from those involved in the EDIT "exemplar group" treatments.

We had realised that we would have to cope with a heterogeneous *institutional landscape,* with widely differing levels of IT capacity. Nevertheless, results from the analysis of institutional IST infrastructures were somewhat frustrating. In spite of the institutions' central role in taxonomic information provision, appropriate infrastructures were often completely lacking, and where in existence, they were rarely outward-looking. Intra-institutional coordination mechanisms were complex, if existing at all. IT developments were mostly depending on soft money funding, with all the consequences for personnel, scope and sustainability. In consequence, we knew that instigating institutional collaboration (the overarching aim of a Network of Excellence project) would present a long-term challenge. It also became clear that we needed a solution independent of database management and operation systems.

We knew that the *taxonomic data domain* was well analysed and to a large extend covered by existing information models and data standards. What was missing was an information model incorporating all these existing, partially overlapping schemes.

It also became clear that good *software applications* were available for taxonomists, for example for descriptive data and the generation of identification keys. Good technologies and an open source environment were available for geographical applications. Handling of bibliographic references was well covered by existing scientific software, including access to citation databases.

Gaias (independent), Marcin Gąsior (MIZPAN), Marc Geoffroy (BGBM), Niels Hoffmann (BGBM), Patricia Kelbert (BGBM), Alexander Kroupa (SMNS, MfN), Dilan Latif (SMNS), Eun-Mok Lee (BGBM), Katja Luther (BGBM), Ôna Maiocco (UPMC), Bart Meganck (RMCA), Dominik Mikiewicz (MIZPAN), Maciej Posluszny (MIZPAN), Francisco Revilla (BGBM), Pere Roca Ristol (CSIC), Pablo Sastre Olmos (CSIC), Bernard Scaife (NHML), Dusan Senko (IBSAS), Lutz Suhrbier (Freie Universität Berlin), Franck Theeten (RMCA), Maxime Venin (UPMC), and Julius Welby (NHML). Exploratory work, information gathering, modelling, and software testing was carried out by Lisa Banfield (RBGE), Franck Dorkelt (INRA), Charles Hussey (NHML), Imre Kilian (HNHM), Boris Jacob (BGBM), Lellani Farina-Crespo (RMCA), Naomi Korn (NHML), Wolf-Henning Kusber (BGBM), Elise Kuntzelmann (UPMC), Barbora Šingliarová (SAVBA), Stanislav Spaniel (SAVBA), David Taylor (RBGK), Dorottya Varsányi (HNHM), Elke Zippel (BGBM), and Magda Zytomska (MIZ PAN). Workpackage coordination was effected by Malte Ebach, Anke Hoffmann, and Agnes Kirchhoff (in that sequence) at the BGBM.

The two most important *data access needs* of taxonomists were tackled by large-scale international initiatives: access to specimen information by the Global Biodiversity Information Facility (GBIF) and access to digitised taxonomic literature by the Biodiversity Heritage Library initiative (BHL). However, an overall integration to cover the needs of taxonomists was lacking. Some of the existing solutions would be difficult to fully integrate, because they depended on specific database or operating systems. Very few solutions existed that supported the full complexity of nomenclatural rules and taxonomic data relations. None was encompassing the full range of data.

Being faced with the unique chance the EDIT project offered, we took the decision to devise, implement, provide and propagate a comprehensive solution for taxonomic computing, the *EDIT Platform for Cybertaxonomy* [1]. The primary objective was to support, enhance and increase the efficiency of the taxonomic work process, for individuals and teams of taxonomists. An explicit aim was to hide the complexity of taxonomic information processing as far as possible, so that it was not inhibiting the workflow, as traditional software applications often did. We knew that new software technologies now offered solutions for some of the problems that had been in the way of creating user-friendly software earlier on. At the same time, the underlying framework had to ensure reusability of the data, seen as the key to future acceleration of taxonomic work processes. On the technical side, hard- and software platform independence had to be ensured to guarantee broad acceptance; at least the newly developed solutions had to be freely available and open source; and for developers wanting to use it for their software projects the solution should provide an API (Application Programming Interface) as well as web services.

In order to achieve these aims, we had to strive to professionalise taxonomic software development. Such a comprehensive solution needed adherence to a strict technological framework. Searching for this framework for development, we looked at content management systems, particularly because using this was a decision taken early-on by another EDIT workpackage -- the “Scratchpads” approach [2]. We saw and see the virtue of this approach for group communication, information dissemination, web publication and aggregation, but we continue to posit that this is not a viable solution for the kind of in-depth treatment of complex data that taxonomists require in their work process. Our aim was principally to support the actual generation of taxonomic data. After weighing several options, Java software development was accepted to provide the most acceptable general framework for Platform application development. Web publication for the Platform can still be realised using content management systems, taking advantage of the Platform’s web services (as demonstrated by the EDIT Data Portal implementations).

3 The Results

Space restrictions allow for only a brief summary of the results achieved so-far. Ciardelli & al. [3] provide a more extensive overview; for full information please refer to the Platform website [4].

The *EDIT Common Data Model (CDM)* now fully covers the data that are

used for systematic treatments resulting from the taxonomic work process (monographs, flora and fauna treatments, and taxonomic checklists). This includes the full complexity of nomenclatural information (botany and zoology), the entire range of taxonomic relationships (including multiple taxonomic hierarchies, synonymies, concept relationships etc.), structured and unstructured descriptive data, geographic information, literature, and specimen data. The CDM is based on existing information models (e.g. the Berlin Model for taxonomic information [5] or the BioCISE model for natural history collections [6]) as well as the standardisation efforts of "Biodiversity Information Standards (TDWG)" -- formerly known as Taxonomic Databases Working Group. Important TDWG standards in this context were the Taxonomic Concept Schema [7], SDD (Structured Descriptive Data) [8], and Access to Biological Collection data [9]. The CDM forms the base for the programming code implemented and made available as the *CDM Programming Library.* The application programming interface or the web services based on the CDM library can be used by programmers to create applications for taxonomists. New functionality created becomes part of the CDM Library after in-depth testing.

As a first step in a user project, a *Community Data Store* is created, i.e. a database that offers the entire scope of information that is covered by the CDM. This can be installed on an individual's computer, on a server in an institutional network, or on servers accessible through the Internet.

Three years of development within EDIT has resulted in a number of CDM-based applications, the two most important of which are the EDITor and the CDM Data Portal.

For data input, the *EDIT Taxonomic Editor (or EDITor)* was developed [10]. It combines an innovative user interface (e.g. allowing full text entry in place of the traditional form-based approach) with the possibility to edit every detail of the database content. The project database can be configured, e.g. by determining which kind of factual data is going to be available for data input (e.g. distributions, threat category, etc.) and which standard terms (if any) are allowed (e.g. TDWG area codes, IUCN threat categories). The taxonomic tree can be displayed and used for navigation and for restructuring by drag and drop. Apart from the taxon-centric standard interface, a "power user interface" presents the data in spreadsheet-like fashion and allows bulk editing and data cleaning. Import and export functionality with several pre-defined formats and standards is implemented. Users can install the EDITor locally on their computer for individual work or access to an institutional Community Data Store, or use it remotely.

The CDM Data Portal is a Drupal-based website used to publish the data in the Community Data Store. It is highly configurable as to displayed content and design. It also offers a taxonomic tree for navigation as well as simple and advanced search functions. The displayed taxon is linked to external resources such as GBIF, BioCASE, BHL, Tropicos, NCBI, Google Images etc. to offer integration with the existing biodiversity information infrastructure. The individual taxon page shows the standard taxonomic data (if the user has configured it that way), i.e. description and factual information. The distribution is visualised through the integrated map viewer (an application of the EDIT Geo-Platform).

All content can be bibliographically referenced. Synonyms can be displayed as homotypic groups, followed by the respective type information. Nomenclatural references are linked to the protologue record (scanned file or web link, where available). An unlimited number of images can be linked and the image gallery offers display in different resolutions and features the image metadata (artist, copyright etc.). CDM Data Portals are in productive use, examples include the EDIT exemplar group sites, for example that for the International Cichorieae Network [11].

Software bundles with the EDITor and Data Portal can be downloaded from the CDM Setup site at http://wp5.e-taxonomy.eu/cdm-setups [12].

Apart from on-line output, functions for pre-formatted print output are being implemented. Out of the (EDIT) box there will be ready made stylesheets for a botanical monograph, a zoological monograph, botanical and zoological checklist, and for the publication of new names in specific journals. Institutional developers will be able to create custom stylesheets conforming to the editorial rules of their in-house publication series.

EDIT has also developed a number of software applications that are not directly CDM based, of which three should at least be mentioned here: (i) The EDIT Geo-Platform [13], [14]; (ii) ViTaL, the Virtual Taxonomic Library, which (in close collaboration with the Biodiversity Heritage Library Europe project) provides an integrated index to taxonomic literature, and (iii) the observation databases and data input tools for the All Species Inventories and Monitoring sites of EDIT workpackage 7.

4 Conclusion

For more than 2 decades there are efforts in joint modelling, standard-building and application development that provide us with excellent knowledge of the taxonomic domain's information structures and business rules. The EDIT Platform is the attempt by European institutions to create a sustainable, collaborative, and comprehensive software solution to increase the efficiency of the taxonomic work process

Acknowledgements

Apart from the EDIT collaborators mentioned in the title page footnote, we would also like to thank numerous taxonomists for their input, in particular those involved with the EDIT exemplar groups: Irina Brake (NHML), Bill Baker, Simon Mayo and Soraya Villalba (RBGK), and Norbert Kilian, Ralf Hand and Eckhard von Raab Straube (BGBM). Gregor Hagedorn gave most valuable advice especially with regard to descriptive data modelling. This work was supported by the European Commission's 6th Framework Programme (Contract No.: 018340).

References

[1] M. Döring and W. G. Berendsohn, "A general concept for the design of the EDIT Platform for Cybertaxonomy", *EDIT newsletter,* vol. 3, pp. 13-15, 2007.

[2] V. S. Smith, S. D. Rycroft, K. T. Harman, B. Scott and D. Roberts. "Scratchpads: a data-publishing framework to build, share and manage information on the diversity of life", *BMC*

Bioinformatics, vol. 10 (Suppl 14): S6doi:10.1186/1471-2105-10-S14-S6, 2009.

[3] P. Ciardelli, P. Kelbert, A. Kohlbecker, N. Hoffmann, A. Güntsch and W. G. Berendsohn, "The EDIT Platform for Cybertaxonomy and the taxonomic workflow: selected Components", *Lecture Notes in Informatics (LNI),* vol. 154, pp. 625-638, 2009.

[4] Anonymous, "EDIT Platform for Cybertaxonomy, "http://wp5.e-taxonomy.eu, 2010.

[5] W. G. Berendsohn, M. Döring, M. Geoffroy, K. Glück, A. Güntsch, A. Hahn, W.-H. Kusber, J. -J. Li, D. Röpert and F. Specht, "The Berlin Taxonomic Information Model", *Schriftenreihe Vegetationsk.,* vol. 39, pp. 15-42, 2003.

[6] W. G. Berendsohn, A. Anagnostopoulos, G. Hagedorn, J. Jakupovic, P. L. Nimis, B. Valdés, A. Güntsch, R. Pankhurst and R. J. White, "A comprehensive reference model for biological collections and surveys", *Taxon,* vol. 48, pp. 511-562, 1999. (Preprint: http://www.bgbm.org/biodivinf/docs/CollectionModel/, accessed 2010).

[7] R. Hyam (Ed.), "Taxonomic Concept Schema – User Guide", Biodiversity Information Standards (TDWG), http://www.tdwg.org/fileadmin/subgroups/tnc/User_Guide.pdf, 2008.

[8] G. Hagedorn, K. Thiele, R. Morris and P. B. Heidorn, "The Structured Descriptive Data (SDD) w3c-xml-schema, version 1.0", Biodiversity Information Standards (TDWG), http://www.tdwg.org/standards/116/, 2005 (accessed 2010).

[9] W. G. Berendsohn (ed.), "Access to Biological Collection Data", Biodiversity Information Standards (TDWG), http://wiki.tdwg.org/ABCD/, 2010.

[10] P. Ciardelli, A. Müller, A. Güntsch and W. G. Berendsohn, "Introducing the EDIT Desktop Taxonomic Editor". In: A. L. Weitzman and L. Belbin (eds.), P*roceedings of TDWG 2008*, Fremantle, Australia, http://www.tdwg.org/proceedings/article/view/325, 2008.

[11] R. Hand, N. Kilian and E. von Raab-Straube (eds.), *International Cichorieae Network: Cichorieae Portal*, http://wp6-cichorieae.e-taxonomy.eu/portal/, 2009+ (continuously updated).

[12] A. Kirchhoff, A. Kohlbecker, N. Hoffmann and A. Güntsch, "CDM setups site - How to install the software modules of the EDIT Platform for Cybertaxonomy", *EDIT Newsletter,* vol. 21, pp. 6-7, 2010.

[13] P. Sastre, P. Roca, J. M. Lobo and EDIT co-workers: "A Geoplatform for improving accessibility to environmental cartography", *J. Biogeogr.*, vol. 36, p. 568, 2009.

[14] P. Mergen and B. Meganck, "Geospatial components for EDIT", *EDIT Newsletter,* vol. 5, pp. 14-17, 2007.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 7-11.
ISBN 978-88-8303-295-0. EUT, 2010.

Descriptive Data in the EDIT Platform for Cybertaxonomy

Maxime Venin, Agnes Kirchhoff, Hélène Fradin, Anton Güntsch, Niels Hoffmann, Andreas Kohlbecker, Elise Kuntzelmann, Ôna Maiocco, Andreas Müller, Régine Vignes Lebbe, Walter G. Berendsohn

Abstract — This paper describes the integration of structured descriptive data in the EDIT platform for Cybertaxonomy. The platform is composed of several software modules supporting the taxonomic workflow from data capture and storage to publication. Descriptive data play an important role within the taxonomic work process. The integration of these data via import/export modules to and from the platform and the publication as natural language output or as keys are explained.

Index Terms — platform, software, taxonomy, description, key, natural language, structured descriptive data, SDD, Common Data Model, EDIT.

1 Introduction

One of the achievements of the European Distributed Institute of Taxonomy (EDIT) [1] is the Internet Platform for Cybertaxonomy, which provides software tools supporting and accelerating the taxonomic workflow (Fig. 1). "A main goal of the Platform is to provide an open architecture to allow connection and integration of existing applications and to provide new developments where necessary" [2]. The Platform is based on the Common Data Model (CDM), which is essentially a description of all data that can be used and edited in the Platform, such as taxon names and concepts, literature references, specimens, distributions, and structured and unstructured descriptive data. All data are stored in a repository known as the CDM Community Store. Different communities can set up their own Store, e.g. to work on a specific monograph, checklist or Flora/Fauna treatment.

The various Platform components are linked by interfaces to the Community Store, for example the Taxonomic Editor (EDITor) for data entry and the EDIT

M. Venin, H. Fradin, E. Kuntzelmann, Ô. Maiocco, and R. Vignes Lebbe are with the Muséum National d'Histoire Naturelle (UPMC-MNHN), CP48, 57 rue Cuvier, 75231 Paris Cedex O5, France, E-mail: venin@mnhn.fr.
A. Kirchhoff, A. Güntsch, N. Hoffmann, A. Kohlbecker, A. Müller, W. G. Berendsohn are with the Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universtität Berlin, Königin Luise Str. 6-8, 14195 Berlin, Germany, Email: a.kirchhoff@bgbm.org.

Data Portal for data publication (see Berendsohn, this volume).

The CDM code Library forms the heart of the Platform software. It enables the individual Platform components to interact. Software developers can use the Library to implement taxonomic software without having to re-create the functionality already developed.

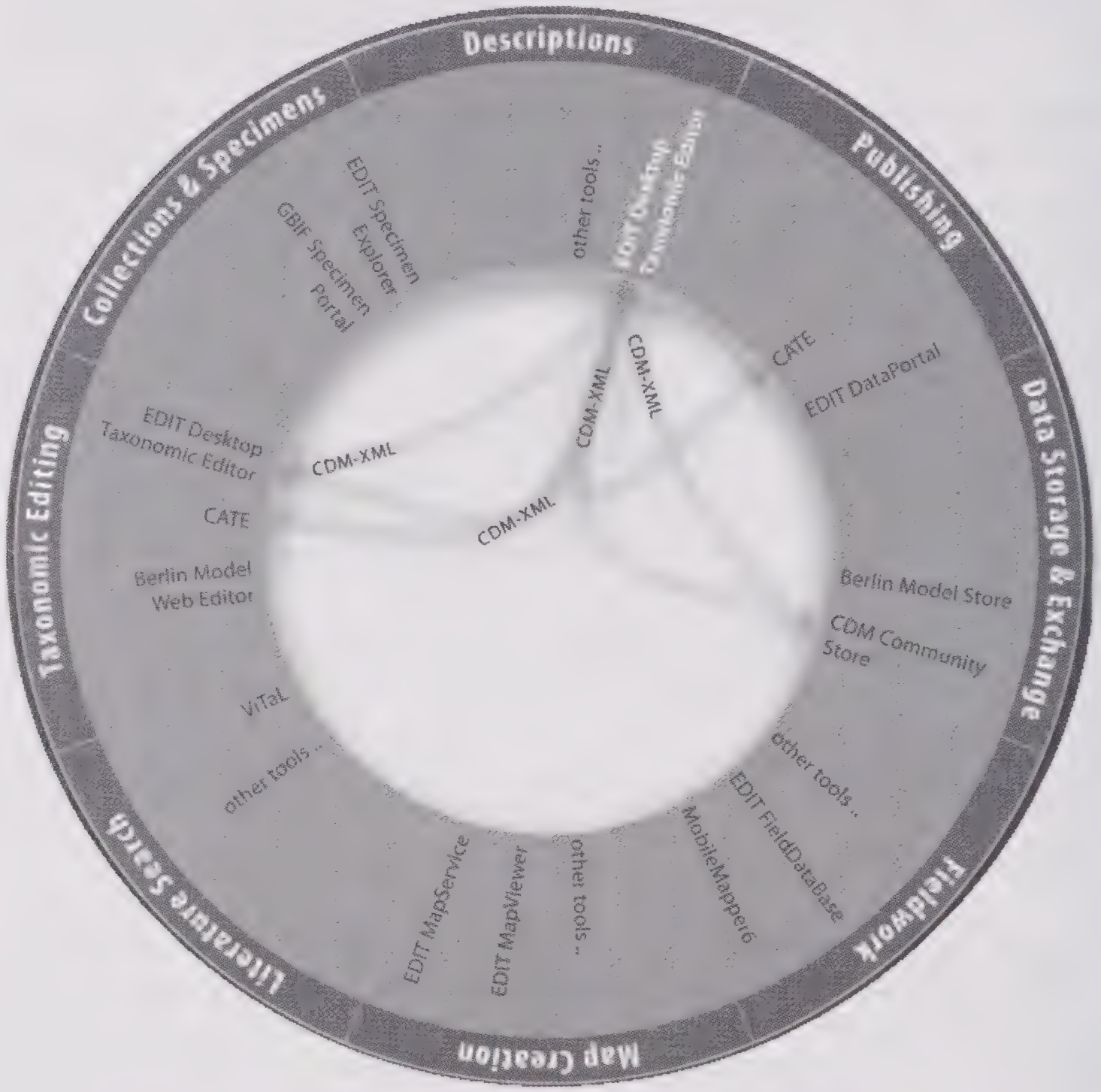


Fig. 1 – Overview over the software modules and functions of the EDIT Platform for Cybertaxonomy (EDIT Cybergate).

The EDIT Platform tools are designed to assist the taxonomist from fieldwork to publication of results, including the management of descriptive data, which play a key role in the taxonomic revision process.

Descriptive data are one of the most important categories of information produced by taxonomists when describing new species or performing taxonomic revisions. Traditionally, taxonomic descriptions were handled as text. However, storing and handling of descriptive data in a highly structured form has strong advantages: data exchange and integration is facilitated, and identification keys (both for printed output and interactive) as well as "natural language descriptions" can be generated automatically and in multiple languages.

There are several established software tools to manage and analyse descriptive data, some of them already existing for decades (e.g. DELTA [3]). Consequently,

it was decided at the outset of the EDIT project not to develop another application but to integrate existing descriptive tools into the EDIT Platform. The key for this is that the CDM complies with the SDD (Structured Descriptive Data) standard [4]. SDD is the current TDWG (Biodiversity Information Standards) standard for descriptive data. Many of the existing descriptive data managing tools (e.g. Lucid [5], Xper² [6], and DiversityDescriptions [7]) do support import and export of SDD conformant data, allowing their users to exchange descriptive data.

2 Structured descriptive data connected to the EDIT platform

2.1 The SDD-CDM Import and Export Module

An SDD-CDM import/export module was developed to integrate descriptive data with the EDIT Platform. Once in the CDM Community Store, the data can be published together with the other information on a specific taxon on the Internet, e.g. by means of the EDIT Data Portal software.

The SDD export from the CDM provides the possibility to use the different software tools and benefit from their specific analytical or output capabilities (interactive identification, comparison of taxa, statistics, etc.) (Fig. 2).

From the technical point of view the general idea behind the import of SDD elements is to create corresponding elements in the CDM in order to allow seamless export/import roundtrips.

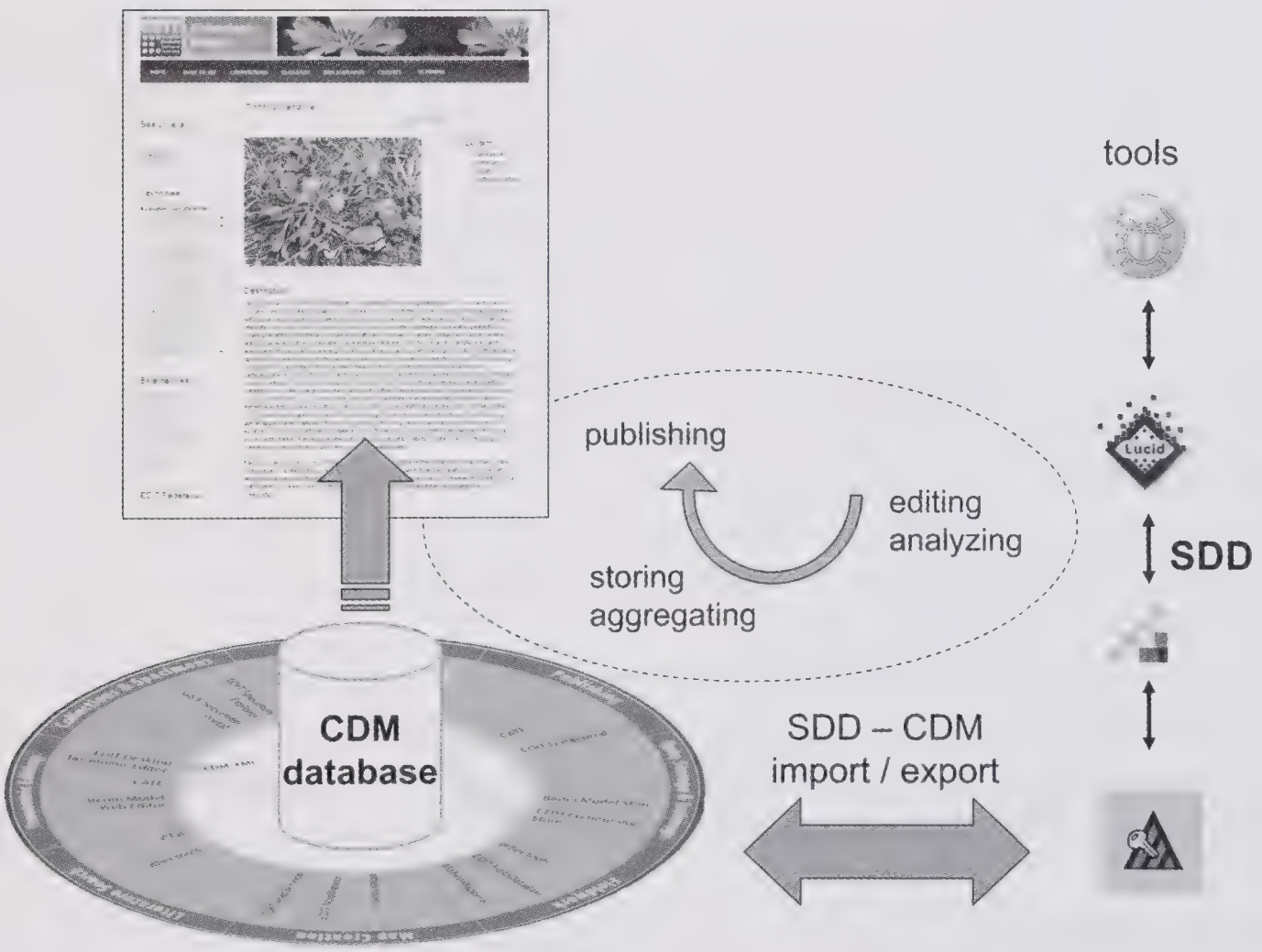


Fig. 2 – Data exchange between descriptive software tools and the EDIT platform for Cybertaxonomy.

2.2 Display of descriptive data as natural language descriptions

The CDM Code Library now includes a feature to generate clear and easy to read output of the descriptive CDM data. The structure of the output can be pre-defined, which allows the scientists to keep a constant scheme, a very helpful feature when preparing output that has to adhere to a defined editorial standard. The natural language description output can be used for publications on the web or for print publications, or simply as a readable preview to control the content of the database.

2.3 Display of keys

A next step in the processing of structured descriptive data is the possibility to automatically generate identification keys from the CDM data. The CDM Code Library now supports output in the form of dichotomous or polytomous identification keys that can be shown on taxon pages of higher taxa in the CDM data portals. Clickable links lead to other key entries or the identified taxa.

The integration of interactive keys will not be possible within the current EDIT project period, for this Platform users have to resort to existing tools. An example of such an integration is given by the CATE project [8], where the LUCID Player is used to provide interactive key functionality on the website.

3 Future developments

The EDIT Taxonomic Editor (EDITor) is the main data entry tool of the EDIT Platform. It allows the editing and presentation of taxonomic information such as classifications, synonyms, taxonomic concepts, descriptions, distributions, specimens and literature references. As any other data in the EDIT Platform this kind of information is stored in the CDM Community Store.

As mentioned above, the EDIT Platform allows choosing among several software tools for the management of descriptive data. One of those is Xper^2, “a management system for storage, editing, analysis and online distribution of descriptive data” [9], which also dynamically creates interactive keys for identifying specimens. This software was chosen as a way forward for the integration of descriptive data into the EDIT platform, because it is Java-based, non-commercial, has been created by an EDIT partner, and it can be integrated with the EDIT Taxonomic Editor.

In the long term, full integration of Xper^2 with the Taxonomic Editor is the aim. A shorter term solution will be to enable Xper^2 to directly work with the data in a CDM Community Store. Xper^2 could then be opened via the EDITor, running as a separate application, but using the same data.

4 Conclusion

With respect to structured descriptive data, the current state of software development for the EDIT Platform for Cybertaxonomy can be summarised as follows:

With the SDD-CDM import/export module the integration of descriptive data into the Common Data Model has been completed.

The natural language module in the CDM library allows users to easily and rapidly generate output describing taxa and specimens. Thanks to the integration with other CDM objects and functions in the CDM Code Library, developers have a very broad range of possibilities to provide users with functions to create, use and publish natural language descriptions.

Generating simple keys is possible with the CDM library. It is an entirely automatic process based on the CDM Community Store. Once the descriptive data have been imported, a taxonomist can directly use this functionality without any extra work.

Acknowledgement

The authors gratefully acknowledge the support of: the EU 6th Network of Excellence Project EDIT (European Distributed Institute of Taxonomy, contract No 018340 - GOCE).

References

[1] N. N., "EDIT - European Distributed Institute of Taxonomy", http://www.e-taxonomy.eu, 2010.

[2] P. Ciardelli, P. Kelbert, A. Kohlbecker, N. Hoffmann, A. Güntsch and W. G. Berendsohn, "The EDIT Platform for Cybertaxonomy and the taxonomic workflow: selected Components", *Lecture Notes in Informatics (LNI),* vol. 154, pp. 625-638, 2009.

[3] M. J. Dallwitz, "A flexible computer program for generating identification keys", *Syst. Zool.*, vol. 23, pp. 50-57, 1974.

[4] G. Hagedorn et al., "The Structured Descriptive Data (SDD) w3c-xml-schema, version 1.1.", *TDWG,* http://wiki.tdwg.org/twiki/bin/view/SDD/Version1dot1, 2006.

[5] N. N., "Lucidcentral", http://www.lucidcentral.org/. *Centre for Biological Information Technology, The University of Queensland, Brisbane,* 2010.

[6] N. N., "Xper^2", http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2. *Laboratoire Informatique & Systématique, Paris*, 2010.

[7] N. N., "DiversityDescriptions", http://www.diversityworkbench.net/Portal/DiversityDescriptions, 2008.

[8] N. N., "The CATE Project", http://www.cate-project.org/, 2010.

[9] V. Ung, G. Dubus, R. Zaragüeta-Bagils and R. Vignes Lebbe, "Xper^2: introducing e-Taxonomy", *Bioinformatics*, vol. 26, no. 5, pp. 703-704, available at http://bioinformatics.oxfordjournals.org/cgi/reprint/btp715v1.pdf, Jan. 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 13-18.
ISBN 978-88-8303-295-0. EUT, 2010.

An online authoring and publishing platform for field guides and identification tools

Gregor Hagedorn, Gisela Weber, Andreas Plank, Mircea Giurgiu, Andrei Homodi, Cornelia Veja, Gerd Schmidt, Pencho Mihnev, Manol Roujinov, Dagmar Triebel, Robert A. Morris, Bernhard Zelazny, Edwin van Spronsen, Peter Schalk, Christian Kittl, Robert Brandner, Stefano Martellos, Pier Luigi Nimis

Abstract — Various implementation approaches are available for digital field guides and identification tools that are created for the web and mobile devices. The architecture of the "biowikifarm" publishing platform and some technical and social advantages of a document- and author-centric approach based on the MediaWiki open source software over custom-developed, database driven software are presented.

Index Terms — field guides, flora, fauna, identification tools, social software, DELTA, SDD, MediaWiki, agile development.

1 Introduction

Digital identification tools may be simple picture guides, printable tabular tools, or interactive tools (single-access, multi-entry, or multi-access keys). A mixture of tools and richly illustrated species pages or glossary definitions is often required. The EU-funded *KeyToNature* project provides a wide spectrum of such tools: together with the "biowikifarm.net", it integrates both the tools and their content. We describe here the architecture and components or this internet-based collaborative authoring and publishing platform.

G. Hagedorn, G. Weber, A. Plank are with the Julius Kühn-Institute, Federal Research Centre for Cultivated Plants, Inst. for Epidemiology and Pathogen Diagnostics, Königin-Luise-Str. 19, D-14195 Berlin, E-mail: gregor.hagedorn@jki.bund.de – M. Giurgiu, A. Homodi, C. Veja are with the Telecomm. Dep., Technical Univ. of Cluj-Napoca, Cluj 400027, Romania – G. Schmidt is with the Institut f. Lern-Innovation, Univ. Erlangen-Nürnberg, D-91052 Erlangen – P. Mihnev, M. Roujinov are with BIKAM Ltd., Sofia 1505, Bulgaria – D. Triebel is with the Center of the Bavarian Natural History Collections, Menzinger Str. 67, D-80638 Munich – R. A. Morris is with the Univ. of Massachusetts, USA – B. Zelazny is with the Internat. Soc. for Pest Information (ISPI) – E. v. Spronsen and P. Schalk are with ETI Bioinformatics, Amsterdam, The Netherlands – C. Kittl, R. Brandner are with evolaris next level GmbH, A-8010 Graz – S. Martellos and P. L. Nimis are with the Department of Life Sciences, Univ. Trieste, I-34127.

2 The MediaWiki Software Architecture

The architecture of the biowikifarm publishing platform is based on the "MediaWiki" open source authoring system [1] that is also used by projects of the Wikimedia Foundation (e. g., the Wikipedias, Wikispecies, Wikisource, or the Commons Media Repository [2]). MediaWiki provides an object oriented document storage model of medium granularity (titled chapters called "pages", rather than whole works). The storage model is akin in many aspects to the currently developed "nosql" database management systems [3] (predating these developments, however, MediaWiki typically uses mysql). Namespaces provided by the storage model allow to re-use the basic model for 1st-class content objects as well as for building objects used in hypertext inclusion. Examples for the latter are media items (binary plus metadata) in the "File" namespace or programming blocks and rich text fragments in the "Template" namespace [4].

The template model provides for flexible schema development. Each template defines a class with freely definable attributes (equivalent to an "entity type"), instances of which can be freely embedded into other objects. Template instances can be hierarchically nested.

The MediaWiki platform is a strong open content and social networking platform. Essential features are the support of the requirements of creative commons licenses (perpetuating licences, tracking contributions and attributing all authors of text and media), a version management and comparison system making changes in a large community transparent to the end user, and a layered development system empowering the community to participate in the functional development of the system.

The latter aspect helps to overcome the discrepancy between user needs and developer actions. Traditional software development requires cycles of planning, use-case and information modelling, piloting, implementation, testing, and rollout, often resulting in slow and inflexible development. Although MediaWiki uses an agile variant of this cycle (involving continuous code integration and live alpha version testing), the php-based core code still suffers from slow development. However, the domain of slow development has been minimized. An event driven extension system provides for an ecosystem of independently developed and tested php-based extensions. Furthermore and highly relevant to the success of MediaWiki projects, the domains limited to developers and server owners are supplemented by further layers (templates, CSS, and JavaScript) that are under the control of the content-editing community:

The templating system enables authors to define and render their own data storage and functional schemata. An unlimited number of templates can be defined, and instances conforming to these schemata (typed and semantically defined fields) can then be inserted in many content objects. Templates are central to the ability of MediaWiki to empower the experts in a given knowledge domain to experiment and achieve information schemata satisfying to their needs. For example, *KeyToNature* defined schemata for media metadata and identification keys. The functionality of templates is limited to prevent detrimental influence on the server, limiting possible malfunctions to those objects that include them.

As a negative point, the templating language has arisen as a unique ad-hoc development, may be difficult to learn, and has no debugging support. Interestingly, this may be a result of social engineering to limit the number of users creating new templates on Wikipedia.

Further layers are the CSS and JavaScript integration. Like templates, these layers are stored as normal MediaWiki objects, profiting from the version control and comparison functionality. Since CSS and JavaScript involve potential security concerns, editing of these layers is limited to content administrators. The community focus of these layers was very positive in the *KeyToNature* project and supported multiple more or less successful approaches to field guides and identification tools.

3 The Biowikifarm

The virtual server is designed as a multi-project platform, enabling the joint administration of a large number of separate wikis. Each wiki can be maintained under its own domain name (owned by partners). Whereas the content administration of each wiki is independent, significant synergies are created by managing multiple MediaWikis on a single "wiki farm".

Presently, the biowikifarm hosts the main *KeyToNature* portal, national *KeyToNature* portals (*pedagogical handbook*, *Offene Naturführer*), the *International Society for Pest Information Wiki*, *LIAS glossary*, *Diversity Workbench*, and the *Deutsche Phytomedizinische Gesellschaft Wiki*.

4 Platform Customization Components

4.1 Media management

The biowikifarm maintains two local media repositories for sharing media between all wikis on the platform. The "OpenMedia" repository is the primary repository for Creative Commons-licensed media. It is supplemented by a "SpecialMedia" repository for media that cannot be openly licensed and are available only under bilateral agreements.

Furthermore, the "Commons" repository with over 7 million images is directly integrated through a web service API. All items from Commons are directly usable as if they were available locally. One problem initially encountered was that the Commons servers may occasionally drop web service requests if overloaded. This could be solved by implementing a license-compatible delayed caching solution (every 10 min. in background).

MediaWiki guarantees the attribution requirements of most Creative Commons licenses by linking media usage to a metadata page containing creators and license information. This page also shows images in a higher resolution. However, displaying this information forces the user to navigate away from the present page. Our own usability studies have shown that users expect an enlarged version of the image without leaving the page context and are confused by the default functionality. A JavaScript based image zooming facility was

therefore added to biowikifarm. The first click on an image will enlarge it in an overlay to the page context, to the maximum extent supported by source image and device resolution. The licensing requirements are fulfilled by presenting a link to creator, copyright, and license information as part of this overlay.

4.2 Metadata and information management

Metadata stored in MediaWiki templates are supported by a customized method. A MediaWiki extension harvests all first level templates (on Wiki pages or inside text files uploaded as attachments) and stores the field-value pairs in a MySQL database for fast access. A web service then provides for queries or recent changes, exposing the data as xml for downstream processing (Fedora Commons, GSearch).

By enabling Semantic MediaWiki (SMW), syntactically defined template schemata are semantically annotated using standard ontologies (Dublin Core, FOAF, SIOC). This allows direct semantic metadata search and inference as well as exposure in the OWL/RDF format. Semantic queries can be embedded, creating dynamical content in wiki pages (outside of metadata, SMW is presently further extensively tested by ISPI).

An embedded Flash application, the MediaIBIS search tool, searches the metadata objects stored in the *KeyToNature* online repository. It has a user-friendly multilingual interface and supports both simple and advanced queries. Details are presented in a separate contribution [5].

4.3 Embeddable identification tools

The platform provides several embeddable identification tools. DELTA datasets can be embedded through NaviKey [6], SDD data through IBIS-ID [7] or Xper^2 [8]. The embedding is achieved through a simple custom "IdentificationTool" extension. To embed a tool, users only need to write a simple statement like: *<IdentificationTool>tool=NaviKey5 config=… NavikeyConfig.xml</IdentificationTool>*. DELTA and SDD data may be placed on wiki pages (rather than in binary attachments) and remain directly editable in plain text mode.

4.4 Wiki-editable single-access identification tools

Single-access keys (i.e. tools with a fixed dicho-/polytomous structure) are implemented in a more direct manner than the multi-access keys. They are based on a combination of templates, CSS and JavaScript, all of which are directly editable by administrators of the wiki (no intervention of server administrators is necessary). On any given wiki page one or several single-access keys can be freely embedded as a structural element in a rich-text layout. The keys provide both a tabular, printable view and an interactive (step-by-step) mode [9]. Details about the wiki key implementation are presented in this volume [10].

5 Sustainability and Scalability

Maximizing sustainability in the face of continuous hardware- and software evolution was a major design priority. Hardware independence can be relatively easily achieved by means of server virtualization, making entire servers easily portable from one physical machine to another. Service is assured by follow-up projects (until 2013) plus a longer-term maintenance pledge of the SNSB IT Center (the SNSB is the government agency for the natural history collections of Bavaria).

Software sustainability is more difficult to achieve. The model of isolated systems maintained in stasis for long periods is not applicable to web software that is dependent on a complex software environment and under permanent threat of malicious attacks. Whereas major publishers achieve permanent redevelopment for their in-house-developments, even mid-sized publishers and software developers have often failed to find the necessary resources. Perhaps the majority of internet offers in biodiversity that were backed by scientific institutions or individuals have therefore ceased to exist. A possible solution is built on three pillars: a) building on a carefully chosen open source software that is supported by a large community with a long-term perspective; b) minimizing project-specific custom developments and partitioning them into small, well documented modules (reducing complexity and the steepness of the learning curve for new developers); c) building the platform to the needs of multiple projects, aggregating available resources and achieving synergies.

We consider the long-term sustainability perspective of MediaWiki to be optimal. It is actively developed, the content of the Wikimedia foundation projects tied to the software makes it highly unlikely that it is abandoned in favour of another project, and version upgrades are always fully automatic (in contrast to some other content management systems that require considerable resources to move from one version to another).

Our own developments are designed to be as modular and layered as possible. They involve small php extensions, a set of templates that can be maintained independent of newer developments, and CSS and JavaScript development. Except for the php extensions, the components are directly editable over the web and can be maintained by a community of users and developers.

An attractive feature of the combination of templates and JavaScripts is their locality to specific documents. The system offers the option to run older identification tools in parallel with newer developments. While this may lessen user experience uniformity, it reduces the analysis and testing requirements for new ideas, enabling agile developments in the future.

Finally, and of great importance to scientific publishing, the principle of locality also applies to content. Scientific knowledge is a stage in a development, no final truth. Opinion may often (yet) matter. Unlike typical databases, the platform assumes no homogenous single truth. Dissenting opinion may be published and outdated knowledge may be retained (adding pointers to updates, etc.). Conventional databases may support dissent (e. g., alternative taxonomic hierarchies), but these expensive solutions are typically limited to a specific aspect. On a wiki platform, any update requires no analysis whether it would

corrupt relational assumptions of older publications – contributing greatly to scalability and sustainability.

6 Conclusion

The MediaWiki-based platform is suitable for the development of collaboratively edited flora and fauna projects. It is powerful, extensible and long-term sustainable. We have successfully implemented a set of native or embedded components. Molecular identification extensions are, however, yet missing. The present platform can be adapted to other purposes in order to create an open source online community of such tools and the scientific interests around them. We welcome further partners to share the platform's use and management.

Acknowledgement

This work was supported by the *KeyToNature* Project, ECP-2006-EDU-410019, in the *e*Content*plus* Programme.

References

[1] MediaWiki software, http://www.mediawiki.org/wiki/MediaWiki, 2010-07.

[2] Wikimedia Foundation Projects: http://wikimediafoundation.org/wiki/Our_projects 2010-07.

[3] MediaWiki Templates, http://www.mediawiki.org/wiki/Templates, 2010-07.

[4] NoSQL databases (overview). http://nosql-database.org/, 2010-07.

[5] M. Giurgiu, A. Homodi, C. Veja, G. Hagedorn and P. L. Nimis, "A search tool for the digital biodiversity resources of *KeyToNature"*. In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 19-24, 2010.

[6] D. Neubacher, and G. Rambold. NaviKey, a Java applet and application for accessing descriptive data coded in DELTA format. http://www.navikey.net. 2005 (onwards), 2010-07.

[7] M. Giurgiu, G. Hagedorn, and A. Homodi, "IBIS-ID, an Adobe FLEX based identification tool for SDD-encoded multi-access keys". *Proc. of TDWG 2009*, 9-13 Nov. 2009, Montpellier, p. 90, 2009.

[8] V. Ung, G. Dubus, R. Zaragüeta-Bagils and R. Vignes Lebbe, "Xper^2: introducing e-taxonomy". *Bioinformatics,* vol. 26 (5), pp. 703-704; see also http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2, 2010.

[9] S. Opitz and G. Hagedorn, "The jKey wiki key player and builder2. *Proc. of TDWG 2009*, 9-13 Nov. 2009, Montpellier, 2009.

[10] G. Hagedorn, B. Press, S. Hetzner, A. Plank, G. Weber, S. von Mering, S. Martellos and P. L. Nimis, "A MediaWiki implementation of single-access keys". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 77-82, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 19-24.
ISBN 978-88-8303-295-0. EUT, 2010.

A search tool for the digital biodiversity resources of *KeyToNature*

Mircea Giurgiu, Andrei Homodi, Cornelia Veja,
Gregor Hagedorn, Pier Luigi Nimis

Abstract — The European *KeyToNature* project has created a framework to gather large amounts of biodiversity-related digital data and metadata from cooperating providers and to make them publicly available via an advanced search engine tool. The novelty of the solution consists in the creation of specific queries for the search web service in the *KeyToNature* digital repository, in the implementation of a communication protocol between the client application and the repository server, and in other original solutions for metadata presentation.

Index Terms — online digital repository, metadata search engine, Rich Internet Application.

1 Introduction

KeyToNature (www.keytonature.eu) is an EU-funded project focusing on interactive educational tools for the identification of organisms. It aims at enhancing the knowledge of biodiversity at all educational levels across Europe. Some project partners are data providers for an online repository for metadata of media resources which can be used in the creation of interactive, computer-aided identification keys. These digital objects should become online searchable and accessible. The solution was to create an online digital object repository that stores only the metadata associated with the digital resources. The associated search tools, based on the assessment of user needs, are described here.

The repository is now searchable via web services, it can interact with other web services which support the management and access of the digital repository [1]. The most important web service is GSearch, which indexes the Fedora

M. Giurgiu, A. Homodi, C. Veja are with the Telecommunications Department, Technical University of Cluj-Napoca, Cluj 400027, Romania, E-mail: Mircea.Giurgiu@com.utcluj.ro.
G. Hagedorn is with the Julius Kühn-Institute, Fed. Research Centre for Cultivated Plants, Königin-Luise-Str. 19, D-14195 Berlin, Germany, E-mail: gregor.hagedorn@jki.bund.de.
P. L. Nimis is with the Department of Life Sciences, University of Trieste, I-34127, Italy, E-mail: nimis@units.it.

digital repository FOXML objects (Fedora Object eXtended Markup Language) and of supports searching this index.

2 Data management and structure

A metadata exchange agreement has been created within *KeyToNature*, so that all metadata in the repository have standardized semantics and support a range of syntactical options for data transfer. The metadata for each resource include information on, e. g., title, creators, keywords, taxa, geographic location, copyright, license, format, access type (printed, offline, online, free, login), URIs, etc. [2]. The actual aggregation of these biodiversity metadata occurs by a "metadata harvesting" process, implemented using specialized tools and methods [3].

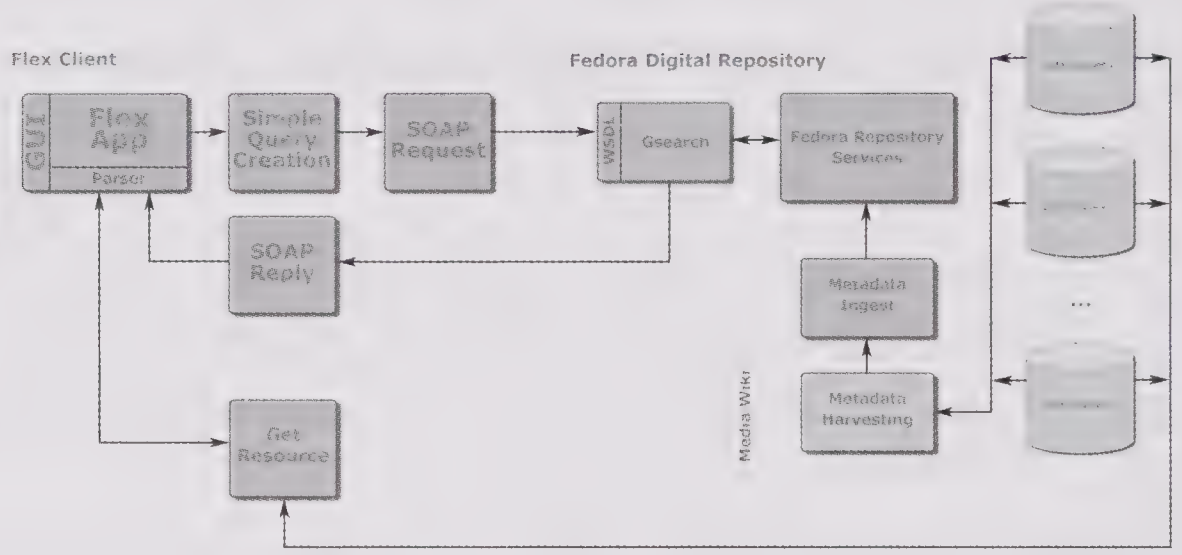


Fig. 1 – The communication between the search client and the digital biodiversity repository of *KeyToNature*.

3 Application design

The chosen framework of Fedora Commons [1] and GSearch is a backend framework (Fig. 1). Although generic search interfaces exists, these are primarily geared towards developers and totally unsuitable for end-users. Furthermore, a goal of the project was to allow search endpoints in various web presences, from the *KeyToNature* portal to various eLearning environments.

The search application was implemented in Adobe Flex [4, 5], which is embeddable in a large variety of web environments. After defining and implementing the communication between the Flex-based client and the Fedora-based digital repository (Fig. 1), the most important step was the creation of the user interface. The interface exposes the methods and mechanisms that the search tool will use in order to transmit the user input (the request or query) to the repository and to present the result for various types of users (beginner to advanced).

3.1 Simple Search

The simple search interface (Fig. 2) consists of a single text input control and additional drop-down menus. The menus allows users to add additional search criteria for narrowing or filtering results. Examples are selections according to the resource type they wish to find (images, identification keys, etc.) or according to availability (online, free, printed-only, etc.).

The exhaustive search for organism names uses a thesaurus of synonyms. This is a complex mechanism that helps users find more resources by extending their search criteria with added synonyms, scientific names and common names. Despite the underlying complexity, the feature is implemented and communicated in the simplest possible way. When the results come back from the repository, users are informed about the extra search terms that were extracted from the thesaurus reply and used in the query.

The simple search interface automatically chooses the best display mode (tabular or matrix image gallery) based on the resource type of the media retrieved.

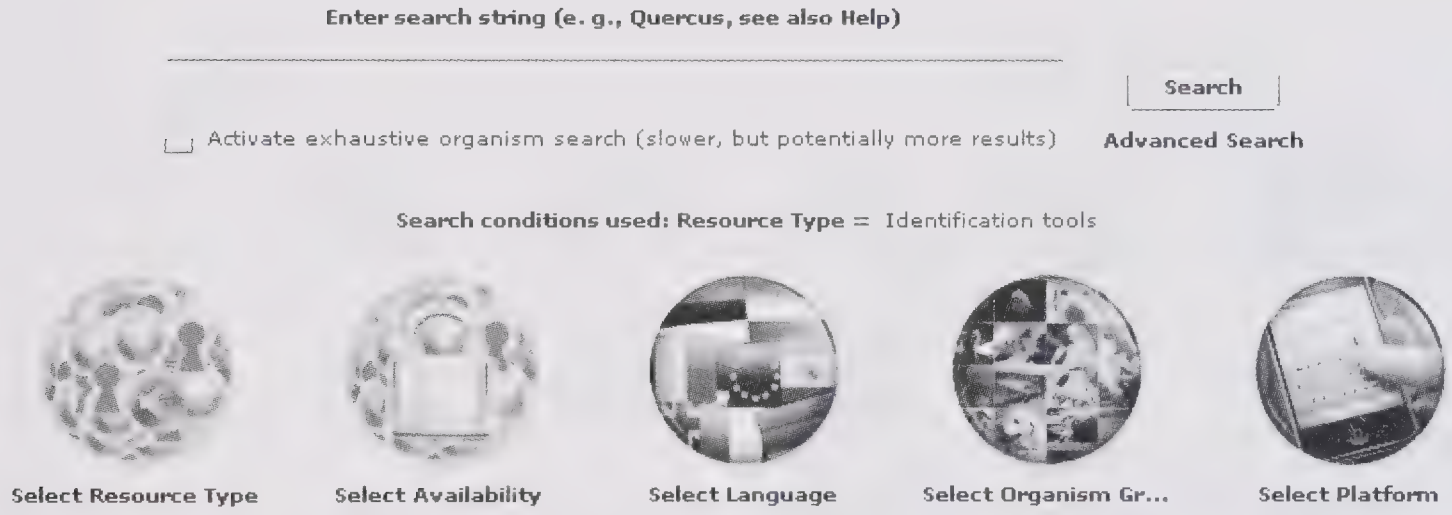

Enter search string (e. g., Quercus, see also Help)
Search
Activate exhaustive organism search (slower, but potentially more results)
Advanced Search
Search conditions used: Resource Type = Identification tools
Select Resource Type
Select Availability
Select Language
Select Organism Gr...
Select Platform

Fig. 2 – The main user interface for simple search.

3.2 Advanced Search

The advanced search interface (Fig. 3) allows users to interactively create complex queries, including logical operators. The interface has three sections: 1) Search Conditions: select the group of searched metadata fields; 2) Sorting Mode: ascending/descending by multiple user-selectable fields; 3) Display mode: a) gallery mode (Fig. 4) with metadata details accessible via the icons situated on the upper right side of the thumbnail image (Fig. 5), or b) table mode.

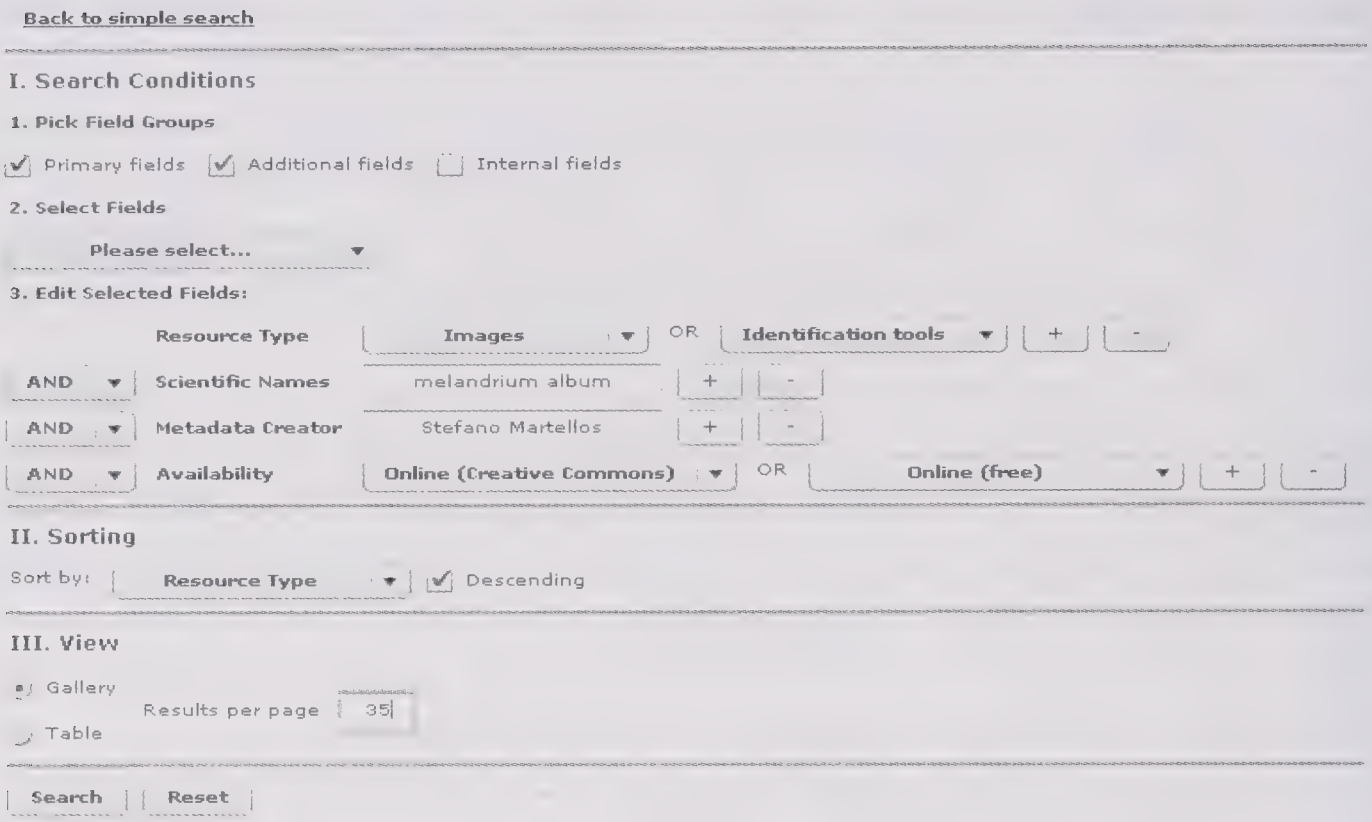


Fig. 3 – The user interface for advanced search.

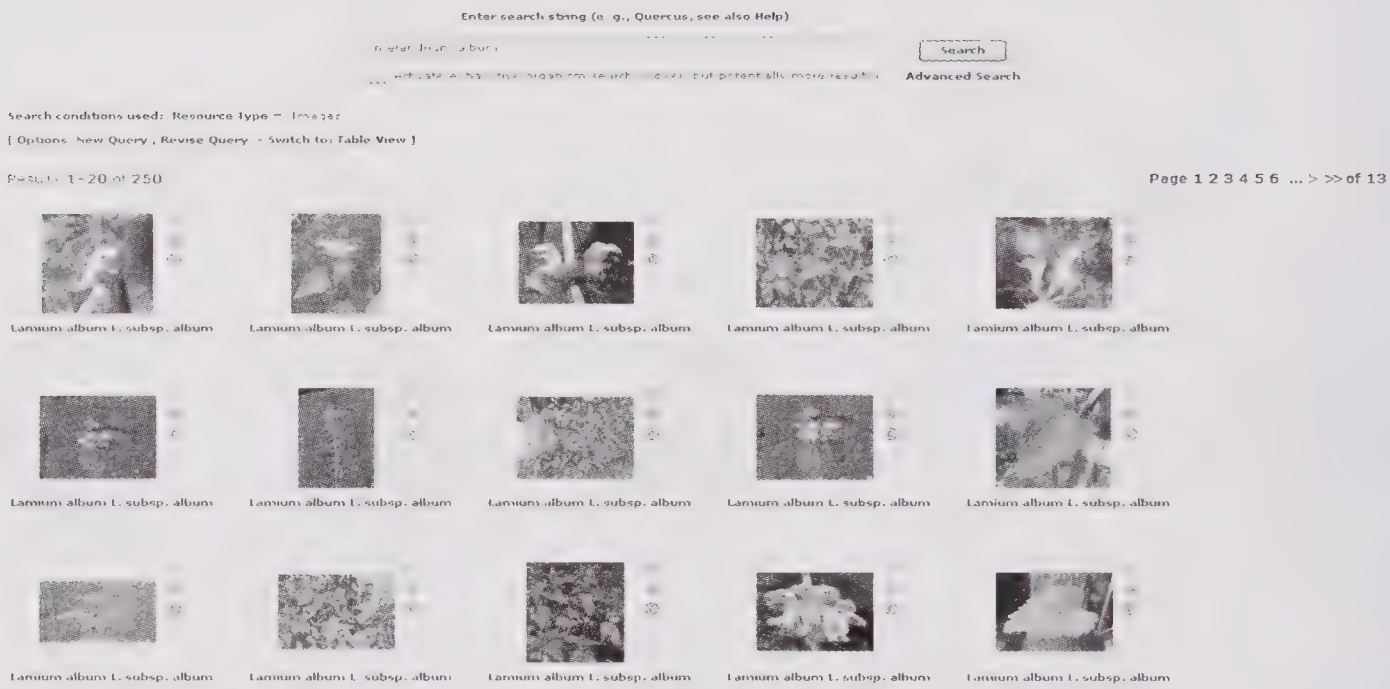

Enter search string (e.g., Quercus, see also Help)
Search
Advanced Search
Search conditions used: Resource Type = Images
[Options: New Query , Revise Query - Switch to: Table View]
Page 1 2 3 4 5 6 ... > >> of 13

Fig. 4 – The search results of digital objects displayed in the “gallery” view mode.

The advanced user interface is based on XML (eXtendable Markup Language) storing the user’s selections and inputs [6], [7]. After a search has been performed, the created query may be easily revised by the user, without having to re-compose a query. When browsing through the displayed results, users have access to three options which are available for both simple and advanced search modes: “New Query”, “Revise Query” and “Switch gallery/table mode”.

3.3 Parameters

The *KeyToNature* search tool currently has several external parameters by means of which it can be preconfigured when embedded into a web interface: 1) language selection (nine languages are presently supported), 2) preset to search identification keys only, 3) preset for searching freely available online

resources only, 5) preset for searching only online resources which are under a "Creative Commons" license.

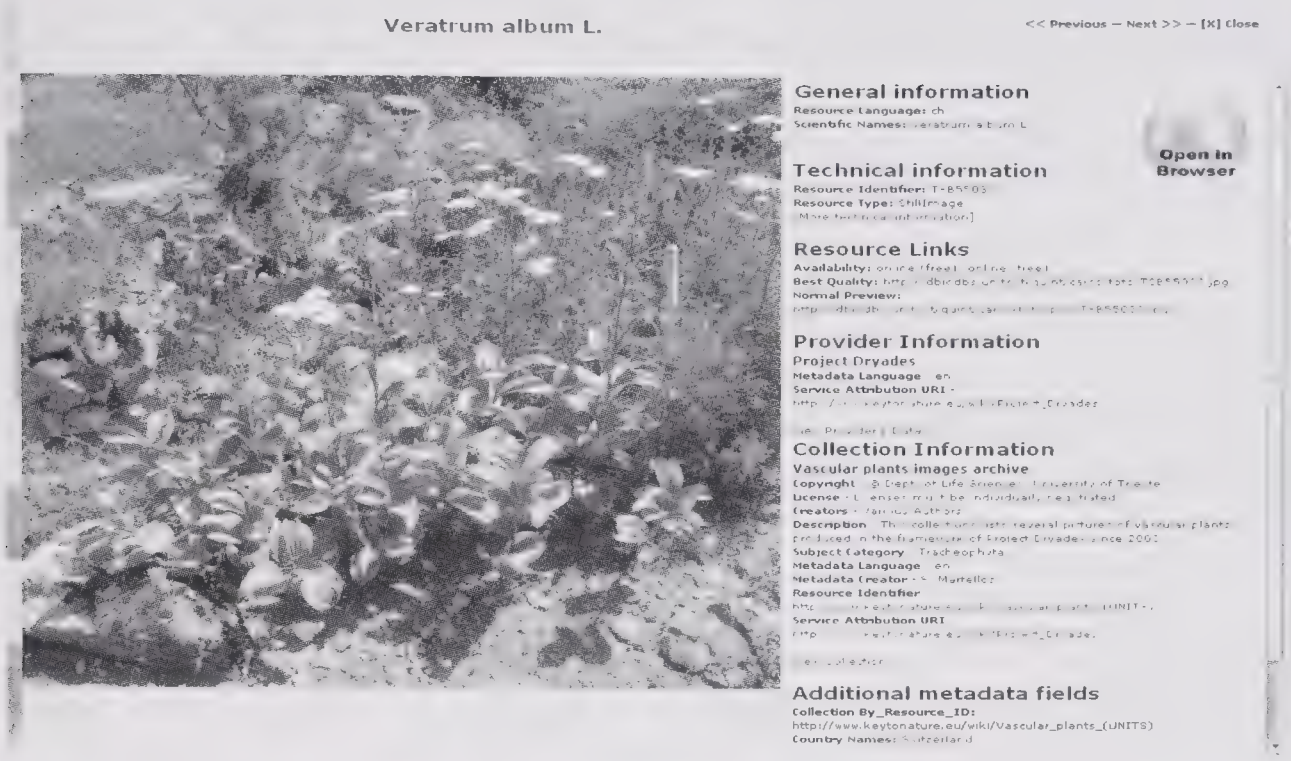


Fig. 5 – Metadata details on a biodiversity-related digital object (in this case, a picture).

3.4 Multilingual interfaces

The *KeyToNature* metadata search engine is currently available in nine languages (Bulgarian, Dutch, English, Estonian, German, Italian, Romanian, Slovenian, Spanish), based on two XML external configuration files, which are easily customizable and extendable to additional languages.

3.5 Testing

The search tool has been carefully tested by project partners as well as experts in software usability. Dedicated wiki [8] pages were created in the *KeyToNature* portal for bug-reporting and suggestions. The reported problems are fixed and suggestions are being analyzed and implemented in an ongoing process.

4 Conclusion

The search tool presented in this paper was implemented as a client application in Adobe Flex. It communicates via specific web services and protocols with the *KeyToNature* online repository of biodiversity-related digital resources. The selection of Adobe Flex has proved a successful decision for implementing the user interface for search. It is an excellent tool for processing the large amount of XML returned by the digital repository. The search application is largely platform-independent, due to the wide availability and distribution of the Flash player. The *KeyToNature* search engine proved to be a robust, fast application, which can be easily integrated into various portals. It could be a model for implementing similar applications interacting with online digital repositories.

Acknowledgement

This work was supported by the *KeyToNature* Project, ECP-2006-EDU-410019, funded in the frame of the *e*Content*plus* Programme.

References

[1] D. Davis and C. Wilper, "Fedora Commons Web Service Interfaces", http://www.fedora-commons.org/confluence/display/FCR30/Web+Service+Interfaces, July 2010.
[2] G. Hagedorn, P. L. Nimis, et al., "Resource Metadata Exchange Agreement", http://www.keytonature.eu/wiki/Metadata_agreement, July 2010.
[3] C. Veja, M. Giurgiu, G. Weber, and G. Hagedorn, "MediaWiki Interoperability Framework for Multimedia Digital Resources", *Proc. Of Int. Conf. on Intelligent Computer Communication and Processing,* 26-28 August 2010, Cluj (Pending Publication).
[4] J. D. Herrington, and E. Kim, *Getting started with Flex 3*, O'Reilly Media Inc., 2008.
[5] C. E. Brown, *The Essential Guide to Flex 3*, Apress, 2008.
[6] E. R. Harold, and W. S. Means, *XML in a nutshell*, O'Reilly Media Inc., 2004.
[7] E. T. Ray, *Learning XML,* O'Reilly Media Inc., 2003.
[8] D. J. Barrett, *MediaWiki*, O'Reilly Media Inc., 2008.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 25-30.
ISBN 978-88-8303-295-0. EUT, 2010.

Developing Web-based Search Portals on Birds for Different Target Groups

Renate Steinmann, Andreas Strasser, Andrea Mulrenin,
Amy Trayler, Sander Pieterse, Ivan Teage, Michael De Giovanni,
John J. Borg, Noel Zammit

Abstract — This paper presents the experiences and interim results from the ongoing iterative development and testing of four distinctive search portals on birds. The search portals are developed within the EU STERNA project and address different target user groups. Based upon specific use case scenarios the search portals are tested and validated in four specific phases, applying three different testing methods: WAMMI online evaluation, focus group evaluation and task-based usability tests. The paper introduces the four search portals, depicts the testing methodology and presents the first results from the ongoing user validation process.

Index Terms — digital library, web based search portals, iterative testing and development.

1 Introduction

This paper presents the ongoing iterative development and testing of four search portals on birds, each addressing a particular target user group. The search portals have been developed as part of the EU funded *e*Content*plus* project STERNA. STERNA is a best practice network and stands for Semantic Web-based Thematic European Reference Network Application (http://www.sterna-net.eu). STERNA comprises 13 organisations and research institutes in the fields of natural history, wildlife and biodiversity. The project started in June 2008 and will finish in November 2010.

R. Steinmann, A. Strasser and A. Mulrenin are with the Salzburg Research Forschungsgesellschaft mbH, Jakob-Haringer-Strasse 5/3, 5020 Salzburg, Austria. E-mail: (renate.steinmann, andreas.strasser, andrea.mulrenin)@salzburgresearch.at.
S. Pieterse is with The Netherlands Centre for Biodiversity Naturalis, PO Box 9517 2300 RA Leiden. E-mail: sander.pieterse@ncbnaturalis.nl.
A.Trayler is with Archipelagos, Institute of Marine Conservation, PO Box 1 Rahes 83301, Ikaria, Greece. E-mail: amy.trayler@archipelago.gr.
I. Teage is with Wildscreen, Bristol BS14HJ, UK. E-Mail: ivan.teage@wildscreen.org.uk.
J. J. Borg, M. De Giovanni and N. Zammit are with Heritage Malta, Valletta VLT03, Malta. E-mail: (john.j.borg, michael.de-giovanni, noel.zammit)@gov.mt.

Following the goals of the European Digital Library Initiative, STERNA seeks to create a distributed and networked information space on nature and wildlife. The main architecture of STERNA utilizes semantic web technologies and standards which allow distributed querying of content based on metadata represented in the RDF (Resource Description Framework) format as well as reference structures represented in the SKOS (Simple Knowledge Organisation System) format.

Related work to this paper includes a short paper depicting the user validation process of STERNA which we have submitted to the Euromed 2010 Conference in Cyprus.

The following chapters describe the four search portals (Chapter 2), the ongoing user validation process and methodology (Chapter 3) and present results that are available after finishing the first two phases of testing (Chapter 4).

2 Four STERNA Search Portals on Birds

Birds and bird-related information are the main focus of the STERNA information space. STERNA addresses a wide range of target users that are interested in birds, ranging from dedicated bird watchers to wildlife enthusiasts in general.

Against this background, we wanted to develop search portals that address different user communities and that meet their specific requirements. To do so, we decided to adopt an iterative design and development process, in which we continuously develop, test and refine different search portals.

The search portals are based on four use case scenarios which were developed by STERNA partners.

The use case of the Netherlands Centre for Biodiversity Naturalis (NCB Naturalis) is targeted at bird watchers, a dedicated group of users that ranges from the more casual to the professional bird watcher. Wildscreen/ARKive, a UK based organisation promoting the appreciation of wildlife, has developed a use case that targets a young and digitally savvy audience. The use case developed by Archipelagos, a non-profit environmental organisation that focuses on the marine and terrestrial biodiversity of the Aegean Sea, addresses boaters and tourists at sea; the fourth use case was developed by Heritage Malta and is targeted at the "humble rambler", a more general audience that is interested in birds.

All four use cases were developed based on a common template and included one or more user scenarios, i.e. typical situations in which users would make use of the STERNA information space to search for bird-related information.

Based on these use cases we developed prototypes of four distinctive search portals. Each search portal addresses a specific target group and offers particular search functionalities: While the Wildscreen/ARKive and Archipelagos search portals provide text based search functionalities, the NCB Naturalis search portal, which is aimed at the more experienced bird watchers, allows users to search for bird related content in different ways and to further refine search results (faceted navigation). In addition to text based search the Heritage Malta search portal also offers a silhouette driven search functionality to better identify and learn about birds.

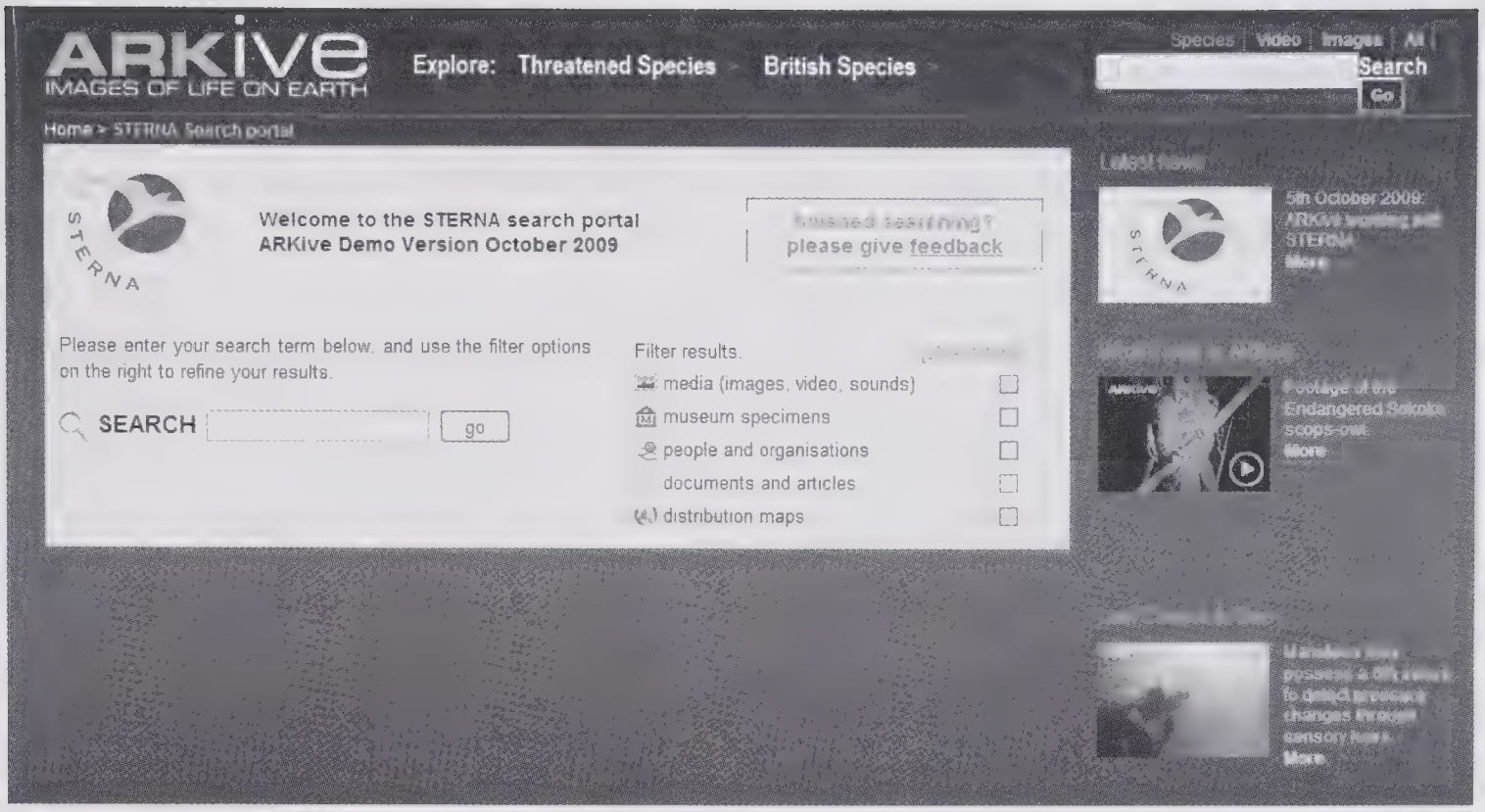


Fig. 1 – Homepage of the Wildscreen/ARKive search portal for young, digitally savvy users that was evaluated with WAMMI.

3 Iterative Development and Testing of STERNA Search Portals

After building prototypes of the four search portals we started iteratively testing and improving them against our target users.

Testing and development in STERNA follows an iterative and user-centred design approach for developing interactive systems. The major principles of this approach are a strong focus on the actual end users—which means that a system should suit the user rather than making users suit the system—and iterative design—which means that the search portals are designed, tested and modified repeatedly. Users are involved actively and early on in the design process: this helps us to gain a deeper understanding of user problems and to avoid taking wrong design paths [1].

The iterative testing and development process in STERNA encompasses four phases: A first online WAMMI evaluation; focus group evaluation; task-based usability tests; and a second online WAMMI evaluation. Due to limited available resources, we apply the WAMMI evaluation only for the NCB Naturalis and the Wildscreen/ARKive search portals. The more in-depth focus groups and task-based usability tests are applied for all four search portals.

The focus of testing in all these phases is both on the usability of the user interfaces as well as the presentation of content and search results.

3.1 WAMMI evaluation

WAMMI stands for Website Analysis and Measurement Inventory and is a web based analysis tool for testing and measuring the user satisfaction of a website or a web based solution. User satisfaction is measured in terms of attractiveness, controllability, efficiency, helpfulness, learnability and the overall global usability. WAMMI requests users to fill in an online questionnaire and to assess a web site or solution. It then compares the user reactions with values generated from

a comprehensive reference database of other tested sites and solutions, thus giving a better understanding of the quality of the tested solution(s).

We developed an online WAMMI questionnaire for evaluating the STERNA search portals which includes the standard 20 WAMMI statements as well as additional questions. It also invited users to comment on the ease-of-use of the search portals and to provide suggestions for how to improve them (see: www.wammi.com).

3.2 Focus Group Evaluation

The main objective of the STERNA focus groups is to identify shortcomings and problems of the four search portals as well as develop ideas and suggestions of how to improve them.

Focus groups are conducted with a small number of representative users. A moderator steers the focus group discussion without discouraging the participants from expressing their thoughts. With focus groups we can identify opinions, attitudes and preferences from participants and gather insights that are sparked by group interaction [1], [2].

3.3 Task-based Usability Tests

With usability tests representative users are requested to perform real-life tasks and thereby to evaluate the usability of the search portals.

While performing the tasks users are requested to comment on any usability problems they encounter ("thinking out loud"). Throughout the tests, users are video-taped for documentation and later analysis (with the documentation kept anonymous and confidential). After the test, the moderator and test users revisit the video documentation of the test in the form of a semi-structured interview, which allows the user to reflect on the test and to provide further feedback [2], [3], [4].

4 First Results of User Testing and Development

4.1 Status of user testing and development

User validation of the Wildscreen/ARKive and the NCB Naturalis search portal started in October and November 2009 respectively with the first round of WAMMI evaluation. Testing continued until early February 2010 when we received WAMMI evaluation reports and content analyses of the user comments provided.

Based on the findings from the WAMMI evaluation, both search portals were improved and then—together with the search portals from Heritage Malta and Archipelagos—tested in focus group evaluations. These took place in June/July 2010. We are currently integrating feedback from focus group evaluations to further improve our search portals. In late July/August, we will conduct the task-based usability tests, to be followed by the second round of WAMMI testing in September/October 2010.

4.2 Findings from WAMMI 1 and focus group evaluations

64 users filled in the online WAMMI questionnaire for evaluating the NCB Naturalis search portal, and 94 users filled in the questionnaire on the Wildscreen/ARKive search portal. Both search portals were rated below average in relation to the WAMMI reference database (i.e. the web sites and solutions that were previously tested), with the Wildscreen/ARKive search portal being rated considerably better than the NCB Naturalis search portal.

The NCB Naturalis search portal received a mean global usability score (GUS) of 21.8, the Wildscreen/ARKive search portal (targeting a young, digitally savvy audience) a mean GUS of 41.4 (on a scale from 1 to 100, where one is lowest and 100 highest; 50 represents the average of the reference database of tested web sites and solutions). For both search portals we received a considerable amount of positive user feedback, which helped us in specifying the main usability problems of the search portals, as well as providing us with valuable suggestions of how to improve them.

The four focus group evaluations conducted in June/July 2010 helped us in further specifying problems of our search portals. After discussing and identifying the main problems of the four search portals, participants provided us with concrete ideas and suggestions for how to tackle these problems and improve the search portals further.

While user feedback from the WAMMI and focus group evaluations was distinctive for each search portal assessed, it also showed us some common problems of our search portals that we need to address.

The visual interface design of the search portals was often not regarded as very attractive and also the search results presented should generally be more visual. Users often remarked that they would like to get more images, video or audio recordings while they were, usually, less interested in metadata lists. Users also noted that the search functionalities and filter mechanisms need to be improved in order to deliver more fitting results to target users. Navigating through the search portal could also be difficult at times, and some users also remarked that they were unsure about the purpose of the search portals.

The search portals thus have to be more intuitive and visually appealing, deliver more fitting results, and their meaning and functionality need to be more apparent for users.

5 Conclusions

The iterative design and testing approach that we applied has helped us in identifying usability problems of our search portals early on in the development process and hence to make the design and development process as resource and cost effective as possible. It has also helped us to better meet the needs and requirements of our respective target user groups. With the next two phases of user testing we expect to further improve our search portals and make them more user-friendly (however, since they are developed as part of a best practice network project, the final search portals will not be “market-ready products”, but advanced prototypes).

References

[1] K. Baxter and C. Courage, *Understanding Your Users: A Practical Guide to User Requirements. Methods, Tools, and Techniques.* Amsterdam, Boston, London, New York, Morgan Kaufman Publishers, 2005.

[2] D. Chisnell and J. Rubin, *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests.* 2nd ed. Indianapolis, Wiley Publishing, Inc., 2008.

[3] B. Albert and T. Tullis, *Measuring the User Experience: Collecting, Analyzing and Presenting Usability Metrics.* Amsterdam, Boston, London, New York, Morgan Kaufman Publishers, 2008.

[4] J. S. Dumas and J. C. Redish, *A Practical Guide to Usability Testing*. 2nd ed. Exeter, Portland, Intellect Ltd., 1999.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 31-36.
ISBN 978-88-8303-295-0. EUT, 2010.

Simple Identification Tools in FishBase

Nicolas Bailly, Rodolfo Reyes Jr., Rachel Atanacio, Rainer Froese

Abstract — Simple identification tools for fish species were included in the FishBase information system from its inception. Early tools made use of the relational model and characters like fin ray meristics. Soon pictures and drawings were added as a further help, similar to a field guide. Later came the computerization of existing dichotomous keys, again in combination with pictures and other information, and the ability to restrict possible species by country, area, or taxonomic group. Today, www.FishBase.org offers four different ways to identify species. This paper describes these tools with their advantages and disadvantages, and suggests various options for further development. It explores the possibility of a holistic and integrated computer-aided strategy.

Index Terms — databases, identification strategy, simple identification tools, web integration.

1 Introduction

FishBase [1] is a Global Species Database (GSD) and a Biodiversity Information System (BIS) on all extant fish species of the world, with about 32,000 valid species currently recognized as valid. It contains a wide range of information and data on biology, ecology, chorology, taxonomy, physiology, human uses, illustrations, etc. and aims at being the web global encyclopaedia on fishes. Four different types of identification tools are made available on the FishBase website (www.fishbase.org), which opens directly the search page:

- 'Eye-balling' drawings and key features by decreasing taxonomic level from class downward;
- Display of all pictures available for a given geographic area or a given family with possible restriction on fin ray meristics;
- Classic dichotomous keys; and
- Uses of simple morphometric ratios.

This paper gives a short description of the tools, how they work and can be used, where the users can find them, their advantages, disadvantages and

N. Bailly, R. Reyes Jr. and R. Atanacio are with the Aquatic Biodiversity Informatics Office, The WorldFish Center and the FishBase Information Research Group, Inc., Los Baños, 4031 Laguna, Philippines. E-mail: n.bailly@cgiar.org.
R. Froese is with the Leibniz Institute of Marine Sciences, IFM-GEOMAR, West Shore Campus, Düsternbrooker Weg 20, D-24105 Kiel, Germany. E-mail: rfroese@ifm-geomar.de.

limitations, and their possible improvements in the next future. Tools not yet in FishBase are also discussed. The need to integrate such tools in a common identification strategy is stressed.

2 Tools

2.1 Quick Identification Tool with outlines

Description, how to use it, and where to find it. - The user is given the choice between boxes displaying simple outline drawings of species representing a given group, together with a short account of key characters. The user clicks on the box, and the boxes for the next level are displayed. For fishes, we start with 6 classes from the most likely ray-finned fishes to the less likely lampreys and hagfishes. Each class leads to its orders, and then each order to its families. Clicking on a family box leads to the "Identification through Pictures" pages described in the next section. It is possible to restrict the identification process to a large geographic area or to a country. This tool is in the top menu in the search page.

Advantages, disadvantages and limitations. - The main strength of the tool is that it is visual. The user is not obliged to read textual accounts. It is quite simple and useful to search for the family. It may be tricky to identify, e.g., eel-like shape groups that can be found in several classes and orders; here, reading the accounts may help. The number of typical species depicted in a box is limited in large groups, and sometimes well-known outlines but rare shapes are not included, such as coelacanths for the lobe-finned fishes box.

Further developments. - Below the family level, the outline drawings are generally not that useful, because species often have the same shape in a given family. There is a possible improvement using the subfamilies for the most species-rich families (Cyprinidae, Characidae, ...).

2.2 Identification through pictures

Description, how to use it, and where to find it. - The principle is to display one typical picture per species by area or taxonomic group. Clicking on a picture opens the corresponding species account with more information.

The tool is accessible from three sections in the search page as "Identification by pictures" under "Information by Family", "Information by Country / Island", and "Information by Ecosystem". For the first section, species within a family are displayed by alphabetical order of scientific names, which shows closely related species next to each other. From that page, it is possible to restrict the search by large area, and/or by the number of dorsal and anal spines. The broad distribution and maximum length of the species are listed as additional help. For the two last sections, the typical pictures of all species reported in the given geographic area are displayed using a traditional sorting of the fish classification, from hagfishes to coelacanths.

Advantages, disadvantages and limitations. - The main strength of this

tool is that it is visual. The user is not obliged to read texts. It is quite simple and useful to search within a family up to 100 species. But identification with pictures only may be difficult or, worse, misleading. As usual, users should read the species account to verify their identification, which is possible by clicking the picture. For groups with over 100 species the tool becomes problematic. Beyond 500 species, the answer time of the server may be prohibitive. Also, only half of the species have at least one usable illustration for identification in FishBase.

Further developments. - A major FishBase goal is to get at least one picture for every species. This is not easily achieved, because many species are only known from a few museum specimens, if any. The tool could be extended to subfamily and genus ranks for the most specie-rich families, and to orders and classes for those with a few species only. For the pages with geographic areas, there should be a taxonomic table of contents/menu at the top of the page, allowing the user to jump directly to the class, order or family (maybe genus) when he knows it.

2.3 Computer-aided dichotomous keys

Description, how to use it, and where to find it. - These are dichotomous keys digitized as they were published in FAO catalogues, revisions, field guides, and major ichthyofaunas. We have developed our own simple database and webpage format for these keys, but we also use LucId Phoenix for an enhanced interface (see [2] for a description). For the FishBase format, couplet numbers, character text, and number of the next couplet, number of the previous couplet, illustration of the character or of the species (+name) stand in 5 columns in a row of an html table (it corresponds more or less to the database table format). Clicking on the number of the next couplet leads to the corresponding row in the table, clicking on the species name leads to the species account. The tool is accessible from the search page, and from the species summary pages under the section “Tools / Identification keys”. Coming from the search page, it is possible to restrict the key selection by large geographic area, Order, or Family, or to enter the key id when this is known. It is possible to specify whether the list should show only keys available with the LucId interface.

Advantages, disadvantages and limitations. - Advantages are the inclusion of pictures in keys that had none, and the ability to easily step forward and backward during the identification process. Also, the species account is only one click away when reaching a possible identification, so that it can be verified with additional information.

Further developments. - The use of the LucId interface is a good development of our simple format. More species pictures and character illustrations are needed. One internal improvement is to be able to give at the same time the name as used in the publication and the current accepted name, highlighting when the name in the publication is now considered as a synonym.

2.4 Identification by Morphometrics Tool

Description, how to use it, and where to find it. - This tool uses measurements that are easy to obtain from specimens or pictures, and computed standard ratios. The user needs to take measurements of Total Length (TL), Head Length (HL), Eye Diameter (ED) and Body Depth (BD). The tool accepts measurements made in centimeters or inches. In the case of pictures, measurements in pixels are accepted. Ratios of the head length, eye diameter and body depth to the total length are computed and compared with values stored in FishBase. Providing the FAO Area, from where the specimen was collected, the Class, and the Family (optional) significantly reduces the number or possible species. In addition, the Total Length (TL), by eliminating species that do not grow as big or larger than the unknown specimen, further focuses the search to a few species in many cases. The tool returns a list of possible species. For each species, a short description, fin counts, a picture and a link to the species summary page in FishBase is included. The "Identification by Morphometrics" can be found in the "Tools" section of the search page.

Advantages, disadvantages and limitations. - This tool can be used when the user does really not know what he has in hands. As far as it was possible, commercial species were covered first. However, such measurements are available "only" for a third of species, limited by the number of suitable pictures from which the reference measurements are taken. It must be also understood that the standard ratios are usually computed on one picture only, so there is no statistical range: the possible range during the matching is predefined.

Further developments. - An advanced interface which includes other measurements such as preorbital length, predorsal length and preanal length is under development. It is expected that including these measurements will shorten the list of possible species most often to 10 species or less.

3 Discussion

Missing tools: interactive keys, image analysis, Barcode. - Since the late 1960s, computer-aided identification softwares have been developed, using a matrix of taxa / character states as the basis of the tool (Delta [3], XPER [4], LucId, etc.). In FishBase, we store data on morphological description of species, but unfortunately in a format that is primarily incompatible with these softwares. A first promising attempt was made to transform the data in a correct format to be included in XPER, which opens the way to the suitable compliance with the TDWG SDD standard (Structured Descriptive Data [5]). But we face the usual content issue that is recurrent for large groups: how to describe a seahorse, a tuna, a hagfish, and a turbot with the same characters? So the limitations are now on the structure and the standardization of descriptions more than a technological issue. Image analysis is another tool that we did not implement. After some tries with both public and commercial products, the result did not seem efficient enough in order to be incorporated in FishBase: at the present stage of these tools, it could help to restrict possibilities only as long as our reference picture collection is more complete than currently; also,

these reference pictures must be prepared “manually” to remove noise (e.g., focusing on the individual and not the background) if we want to increase the true positive matching. The other issue is to find solutions to use at the same time live underwater pictures with different orientations of the individual and dead specimens with different colours, not to speak about growth allometries, sexual dimorphism, and intermediary individuals when their sex changes during their life cycle like in some families. A new approach is the identification offered by the Barcode of Life (BoLD website [6]), if the user has access to a respective test kit for the genetic analyses. It is not clear whether this identification tool can be more integrated in other webpages like FishBase (the two websites are already cross-linked), or if it is best to use the tool from the BoLD website.

Identification strategy. - At the moment, identification tools in FishBase are independent, except that the Quick Identification Tool links in the end with the Identification Through Pictures Tool beyond the family level. The idea is that all the tools should be integrated into one, and that the user could choose to jump from one to another anytime during an identification session, or even better, that the system could guide the user in that choice, according to the declared skill level of the user. Interactive keys are obviously the start of such developments. XPER proposes already some modules that can build the identification pathway according to some constraints (e.g., sort out species first that are the most abundant, the easiest to identify, or have the most striking forms): someone has to enter the relevant data. But we are in need of both the design and the technology to jump from outline drawings to identification key and to real pictures, from morphometrics/meristics to image analysis, restrict to a size or a geographical area. The final vision is that the system would guide the user and suggest how to start and which pathway to use across all tools.

4 Conclusion

FishBase has deliberately favoured the simplicity over more elaborated identification tools that are costly to develop and maintain, including in terms of data. Some of these simple tools are really easy to deploy on the web such as the Quick Identification Tool with simple outline drawings, and the computerization of the printed dichotomous keys under a simple database format and web layout. Colleagues could move forward quickly to these simple solutions for other taxa. However, each of the tools existing or not in FishBase is interesting in a given context. A long-term goal for the computer-aided identification domain could be to gather all tools in a unique strategy and interface for the user. But this still requires research and technological development, such as the work being done under the European project *KeyToNature* [7]. The last important point is that illustrations, including correctly identified pictures, must be made publicly available. Images are used in 3 of our 4 tools, and we could design the morphometrics tool only because we had a significant number of pictures available. *Homo sapiens* is a species that uses the visual sense to a high degree, and visual identification is still and may remain for a long time its preferred method.

Acknowledgement

The various tools in FishBase were developed in the last 20 years during various projects mainly funded by the European Commission. Eli Agbayani, Josephine Barile, Elijah Laxamana, Christian Elloran and Stacy Militante were the successive programmers who developed them.

References

[1] R. Froese and D. Pauly (eds.), *FishBase 2000: Concepts, design and data sources*. Los Baños, Philippines, ICLARM, xvii+344 pp., 2000.

[2] Lucid Central, "Lucid Phoenix", http://www.lucidcentral.org/Software/LucidPhoenix/tabid/152/Default.aspx, 2010.

[3] M. J. Dallwitz, "Overview of the DELTA System", http://delta-intkey.com/www/overview.htm, 2009.

[4] J. Lebbe and R. Vignes Lebbe, "Xper^2". http://lis-upmc.snv.jussieu.fr/lis/?q=ressources/logiciels/xper2. 2010.

[5] G. Hagedorn, K. Thiele, R. Morris and P. B. Heidorn, The Structured Descriptive Data (SDD) w3c-xml-schema, version 1.0. http://www.tdwg.org/standards/116/. [Last retrieved 05-May-2007], 2005.

[6] BoLD, "Barcode of Life Data Systems" http://www.boldsystems.org/views/login.php, 2010.

[7] *KeyToNature*, "*KeyToNature*: a new e-way to discover biodiversity", http://www.keytonature.eu/wiki/, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
ù Biodiversity: Progress and Problems – pp. 37-42.
ISBN 978-88-8303-295-0. EUT, 2010.

The Catalogue of Life: towards an integrative taxonomic backbone for biodiversity

Frank A. Bisby, Yuri R. Roskov

Abstract — The Catalogue of Life Programme is addresseing the need for a comprehensive catalogue of the world's presently known animals, plants, fungi and micro-organisms. The need is for an electronic catalogue that can be used as a taxonomic back-bone. In a wide variety of programmes covering species and documenting many types of biotic materials and records. The First Phase of the programme has used an architecture based on an array of global species databases to reach coverage of about two-thirds of known species. In the Second Phase of the Programme there will be a new architecture, a new array of services, and a ring of partnerships with global programmes.

Index Terms — Catalogue of Life, global taxonomic framework, species checklist, synonymic indexing, taxonomic hierarchy.

1 Introduction

Despite 250 years of effort in the taxonomic profession, there is still, in 2010, no complete catalogue of all presently known animals, plants, fungi and micro-organisms of the world. This is a critical problem for the scientific community, and for national, regional and global organisations that organise and regulate the exchange of biotic information and materials worldwide. The set of organisms known to science is a key dimension of human knowledge concerning global biodiversity, evolution, ecology, natural resources, and biotic response to climate change. It supplies a vital set of index terms needed to access most biodiversity knowledge. There is increasing public need and expectation, focussed through the UN Convention on Biological Diversity (CBD), to complete such a catalogue of all known organisms for international uses. Many commentators are surprised that a complete catalogue does not already exist. In fact it is a non-trivial task that is too large for the individual

F. A. Bisby is with the Species 2000 Secretariat, Centre for Plant Diversity & Systematics, School of Biological Sciences, University of Reading, READING, RG6 6AS, UK.
E-mail: f.a.bisby@reading.ac.uk.
Y. R. Roskov is with the Species 2000 Secretariat, Centre for Plant Diversity & Systematics, School of Biological Sciences, University of Reading, READING, RG6 6AS, UK. E-mail: y.roskov@reading.ac.uk.

capabilities of even the largest taxonomic institutions, due to the distributed nature of the knowledge.

The Species 2000 programme, working in partnership with ITIS in N. America, has made substantial progress with resolving this problem. It has created, maintained and enlarged its Catalogue of Life to the point where it now covers 1.25 million species of plants, animals, fungi and micro-organisms, more than two-thirds of the anticipated total of 1.9 million presently known species worldwide. It has done this by employing a radical architecture of federating global sectors of taxonomic expert knowledge from a growing array of supplier databases, and integrating these into a single taxonomic hierarchy and species checklist. The distributed system harvests taxonomic knowledge provided and maintained by a community of supplier organisations in the taxonomic profession, combining work by the major taxonomic institutions with that of smaller networks and individuals. This process was brought to production scale by the EC EuroCat project funded as a scientific infrastructure under FP 5 (2003 – 2006) and further developed since then with funding from other sources.

Over the last two years the programme has concentrated on extending and improving the scientific content of the Catalogue of Life, which is now a unique and scientifically valuable resource. However, it has come as a bonus to see the rising and now substantial public usage in Europe and all over the world, including by GBIF and the Encyclopedia of Life, of what is presently an incomplete service. The 4D4Life Project provides us with a timely opportunity to develop a parallel focus on services. It will enable us to enrich the variety and technical sophistication of taxonomic services that are undoubtedly possible, exploiting the taxonomic resource that we are already building. The utility of these services will secure the sustainability of the whole programme into the future.

2 The present concept

The Species 2000 & ITIS Catalogue of Life (henceforward 'the Catalogue') has a single purpose, to enable users throughout biological and biodiversity sciences, and across the many scientific and non-scientific disciplines that use organism information, to access data about all organisms by means of a species checklist and a taxonomic hierarchy. It is already used to access data such as organism relationships, ecology, DNA sequences, protection status, invasive properties or information in any one of a myriad of other data domains. Such a Catalogue needs to be:

i) comprehensive: covering all known organisms in all groups;

ii) global: organisms of the whole world, in terrestrial, freshwater and marine environments;

iii) validated: a responsible, modern and professional globalised taxonomic view of the classification, supported by and embedded in the profession's activities;

iv) accurate: reflecting as accurately as is practical the detail of diversity of living organisms;

v) accessible to all: a clear view of the taxonomy, eventually in multi-lingual

presentation;
vi) available to all: widely and freely available in a variety of forms; and
vii) dynamic: updated for taxonomic changes though time, either continuously or annually.

To be effective in the many applications in which it is used, the classification and the naming of species and higher taxa must be as close to 'agreed and correct' as is possible in taxonomy. This means for each taxon either using a consensus system, or selecting by peer review and using consistently one of the competing classifications where alternatives are in wide use. Because alternative classifications have been used both today and in the past, users must be able to locate species known by other names (or concepts) in the Catalogue, and discover alternative names under which to access data on the internet or in other resources. Consequently synonymy and common names must be included for each species. As much as possible should be 'concept-based', a precision provided by some of the supplier databases.

The dream is simple - to create a Catalogue that contains an accurately maintained synonymic species checklist covering all known groups, connected in a validated taxonomic hierarchy.

3 The existing (Phase 1) Programme

The present Catalogue of Life Programme, led by the global Species 2000 organisation based in Reading, and working with the N. American organisation ITIS, was set up as an international programme at a UK-funded (BBSRC) workshop in 2001. Bringing the programme up to production scale was funded by the EC as one of its scientific infrastructures (EuroCat), with further funding by the Japanese Government, the US Government (through ITIS) and GBIF. Output is via the *Catalogue of Life Annual Checklists* on the web [1], and on free DVD [2], and the *Dynamic Checklist* on the web, both also available as web-services for electronic use.

In March 2007 an EC-funded 'Million Species Day' symposium was held to celebrate reaching one million species. The *2010 Annual Checklist* now provides a quality species checklist of 1,257,735 species with unique identifiers and a hierarchy for all organisms (animals, plants, fungi, chromista, protozoa, bacteria, archaea, viruses). The estimate for the number of known extant species is currently 1.9 million [3].

The present Catalogue benefits from simplicity of structure incorporating minimal but standardised data for each species. These contribute to its success in providing a universal baseline needed by all biologists, and in making the project practicable. It consists of two knowledge structures, and software that enables the user to search or traverse them, and to toggle between them. i) The Species Checklist is a series of Species pages that are located by name searches, with automated synonymic indexing. Each page gives the Standard data for a Species, including common names, the higher taxa it belongs to in the hierarchy, and geographical distribution. ii) The Taxonomic Hierarchy is an expansible tree that can be followed down through the classification to the 1.25 million individual species. Or it can be used to navigate upwards to the higher

taxa containing the one that is viewed. By clicking on a higher taxon listed on a species page, the user can transfer to the tree for that taxon, and see all its daughters. Conversely, by clicking on a species at a twig in the tree, the user can visit the relevant Species page in the Checklist.

A comprehensive checklist cannot be made simply by adding together regional or single-country lists. Different classification and naming schemes mean that a simple additive list would be massively duplicative and of little use. The current system is a successful development of the original BBSRC SPICE project. It federates the taxonomic sector checklists provided by a distributed array of global species databases (GSDs), which are globalised checklists of a whole taxon, harvested across the Internet, and fitted together 'end-to-end' within a single overall classification. When enough sectors are fitted this process can eventually create a complete list. The number of GSDs contributing one or more taxonomic sectors to the Catalogue reached 77 for the 2010 edition, including 47 based in Europe, 18 in the USA, 5 in Brazil, and 7 in New Zealand, Russia, Japan, Taiwan, Australia and the Philippines. The model ensures that sectors are enhanced taxonomically by the supplier databases, and ca. 3,000 experts globally contribute to these databases. The whole programme depends on the integration and aggregation of expert knowledge from these key suppliers.

Each GSD sector is attached at its 'top point' (its highest ranking taxon) in the hierarchy, and in addition to harvesting the checklist, the system also harvests branches of the tree beneath this top point for the hierarchy leading down to the species in that sector of checklist. The checklist and hierarchy created from a growing array of GSDs in this present-day architecture ('Architecture 1') referred to as the 'Global Hub'.

Despite the evident success of Architecture 1 in permitting the rapid build up of the Catalogue to its present point, its limitations have been evident for some time. The difficulty is simply that no-one anywhere in the world is creating global species databases for some of the least known taxonomic groups, so by this model these would be destined always to remain as gaps in the Catalogue. In the EC EuroCat project (2003 – 2006), we additionally experimented with making a Regional or 'Euro-hub' with a further set of European regional databases, and versions of SPICE that could handle multiple hubs, and the first steps towards integrating their contents using the LITCHI 2 taxonomically intelligent integrity tracking. We then started to plan an 'Architecture 2', in which an array of Regional Hubs might be connected to the Global Hub, this providing linkage to regional databases from many parts of the world, but also the potential for the Global Hub to harvest data or checklist sectors from Regional treatments for the species groups that were missing from the Global Hub. Good progress is being made with initiating these Regional hubs now, and plans in the 4D4Life Project are to develop a unified concept and specification for this Multi-Hub Network working with the designated centres for China, New Zealand, Brazil, and Australia.

4 The Phase 2 Programme

In June 2009 Species 2000 and ITIS launched the Phase 2 programme of the Catalogue of Life with a fresh funding initiative and extended partnerships around the world planned for the 5-year period 2009 – 2014. In outline Phase 2 involves:

1. A new array of electronic and other services
2. A new service-based cyber-infrastructure: an ecosystem of services
3. A strategy for completing taxonomic coverage of the Catalogue
4. A world-wide multi-hub network of regional hubs
5. A 2nd Edition Catalogue of Life Management Hierarchy
6. A ring of partnerships with global biodiversity programmes

5 4D4LIFE Project

The 4D4Life Project in the EC e-Infrastructure programme has now taken responsibility for the array of new services, the new cyber-infrastructure, and designing the world-wide multi-hub network. Sara Oldfield at Botanic Gardens Conservation International is co-ordinating the Services Team, and Alex Hardisty at Cardiff University is co-ordinating the System Design Team.

6 I4LIFE Project

The i4Life Project in the EC e-Infrastructures Programme will shortly take responsibility for the ring of partnerships with global biodiversity programmes intended to harmonise and integrate between the taxonomic catalogues.

7 Conclusion

Substantial progress has been made with developing a comprehensive Catalogue of Life. However, there remains much to be done in the ambitious Species 2000 & ITIS Catalogue of Life Programme. The Catalogue is still far from complete in terms of taxonomic groups and known species; there is much to be done in improving both quality and fill of the Standard data set across all taxa; the new public services need to be fully tested and rolled out, and the programme needs to make progress with becoming sustainable as a scientific infrastructure for use around the world.

Acknowledgement

This work was supported in part by the EC DG INFSO FP7 e-Infrastructures Programme under the 4D4Life Project (Grant 238988).

References

[1] F. A. Bisby, Y. R. Roskov, T. M. Orrell, D. Nicolson, L. E. Paglinawan, N. Bailly, P. M. Kirk, T. Bourgoin and G. Baillargeon, *Species 2000 & ITIS Catalogue of Life: 2010 Annual Checklist*, www.catalogueoflife/annual-checklist/2010, Species 2000, Reading, 2010.

[2] F. A. Bisby, Y. R. Roskov, T. M. Orrell, D. Nicolson, L. E. Paglinawan, N. Bailly, P. M. Kirk, T. Bourgoin and G. Baillargeon, *Species 2000 & ITIS Catalogue of Life: 2010 Annual Checklist*, DVD, Species 2000, Reading, 2010.

[3] A. D. Chapman, *Numbers of Living Species in Australia and the World*, 2nd Edition. Australian Biological Resources Study, Australian Government, Canberra, 2009.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 43-48.
ISBN 978-88-8303-295-0. EUT, 2010.

BHL-EUROPE: Biodiversity Heritage Library for Europe

Jana Hoffmann, Henning Scholz

Abstract — The Biodiversity Heritage Library for Europe (BHL-Europe) is an EU funded project making available biodiversity literature over various platforms, e.g. a multilingual portal for search and retrieval, the Global References Index to Biodiversity (GRIB), and Europeana. BHL-Europe brings together European digital biodiversity content already available and will assist in future scanning activities. BHL-Europe is asking the scientific community to contribute to the improvement of the BHL-Europe portal functionalities by giving feedback and participating in BHL-Europe workshops or to provide information about unexplored repositories of digital biodiversity content.

Index Terms — Biodiversity Heritage Library for Europe (BHL-Europe), digital library, biodiversity literature, taxonomy, taxonomic intelligence, optical character recognition (OCR).

1 Introduction

The lack of access to the published biodiversity literature is still a challenge in the day-to-day business of taxonomists or researchers dealing with biodiversity-related questions. In the past, only libraries of large and renowned institutions such as universities, natural history museums or botanical gardens housed specific literature indispensable for taxonomic work. Collecting relevant literature on a certain group of organisms was time consuming, cost-intensive and required loans of books or even a visit to the respective institution. Today, quick and easy access to digital literature is more and more important to facilitate scientific work. However, digitisation of literature is expensive and requires a lot of additional work on making the content available for extensive search and retrieval. Furthermore, there are still major problems with right holders, thus limiting the range of content freely available on the internet. For scientists it is of high importance to have a sustainable infrastructure they can rely on with a simple and quick mechanism to search for bibliographic information and free access to digital content of high quality. This is especially true for scientists working in developing countries with limited access to literature in general. As taxonomy is an 'accumulative' science it relies more than other

The authors are with the Museum für Naturkunde, Leibniz Institute for Research on Evolution and Biodiversity at the Humboldt University Berlin, Invalidenstr. 43, 10115 Berlin, Germany. E-mail: jana.hoffmann@mfn-berlin.de, henning.scholz@mfn-berlin.de.

disciplines on a complete record of literature on a group of organisms of interest and has a stronger focus on historic publications. Moreover global availability of digital content of biodiversity literature is also important for training students and early career scientists and helps promoting the importance of taxonomy as a discipline. Additionally, enhancement of availability of biodiversity literature for a wider public raises awareness of the importance of protecting our planets biodiversity.

Over the last few years a large number of library resources for taxonomists have been made available online - including virtual libraries and search engines as well as digital libraries. Since 2007, numerous libraries in the UK and USA are digitising their holdings of biodiversity literature and making them available on the internet. Today, BHL - the Biodiversity Heritage Library (http://www.biodiversitylibrary.org) is certainly the largest digital library for taxonomists offering free access to more than 30 million pages of historical biodiversity literature (as of July 2010) via the internet originating from 12 major natural history museum libraries.

2 Overview of BHL-Europe

In 2009 the European Commission launched a new project 'BHL-Europe - Biodiversity Heritage Library for Europe' (http://www.bhl-europe.eu) within the framework of the *e*Content*plus* program. This project will run for 36 months through April 2012. The BHL-Europe consortium consists of 28 partner institutions (natural history museums, botanical gardens, libraries, right holders and companies) including 26 European institutions and two American institutions representing BHL (US). BHL-Europe aims among others at (1) supporting existing digitisation initiatives with best practice guides, for example, (2) facilitate and enable the initiation of new scanning initiatives, and (3) bringing together existing digital content scattered all over Europe in a number of libraries and natural history institutions. Currently, 18 out of the 28 consortium partners of BHL-Europe are active contributors to the corpus of digital resources. This corpus of more than 100,000 monographs and serial volumes in April 2012 will eventually be available on three platforms (Fig. 1) (1) a multilingual BHL-Europe-Portal for search and retrieval for scientists and public users, (2) the Global References Index to Biodiversity (GRIB), and (3) Europeana.

The technical architecture of BHL-Europe is based on the Open Archival Information System (OAIS) reference model. It is the backend of the multilingual portal for managing content ingestion, archival and delivery of the digital objects (Fig. 2). A prototype of the new portal will be available in fall 2010, but the final system is expected for the end of the project in April 2012.

BHL-Europe and EDIT (http://wp5.e-taxonomy.eu/) are building the Global References Index to Biodiversity (GRIB), a database generated from the partner libraries catalogues and completed with content management and deduplication functionalities, that eventually refers to all of the worlds published biodiversity literature. This will enhance the possibilities of search and retrieval of digital literature for taxonomists significantly. It will also assist librarians in the process of scanning planning. A GRIB prototype is working already and the final system is expected to be finished in spring 2011.

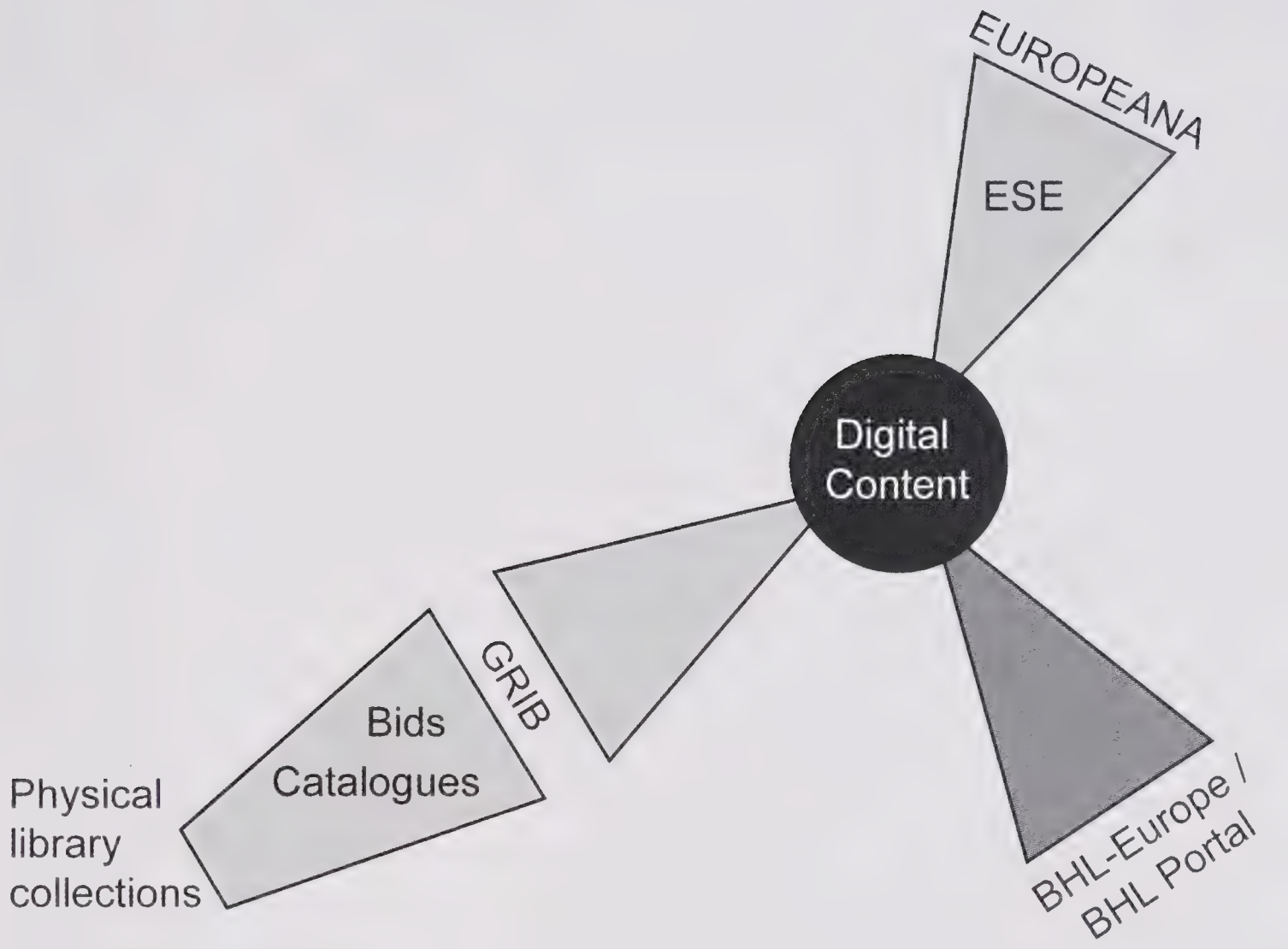


Fig. 1 – The BHL-Europe users (taxonomists, general public) will mostly access the content either through the BHL-Europe / BHL Portal or Europeana (ESE = Europeana Semantic Elements). The major access route for the librarians managing the scanning process is the Global References Index to Biodiversity. It is composed of the catalogue records of the physical library collections.

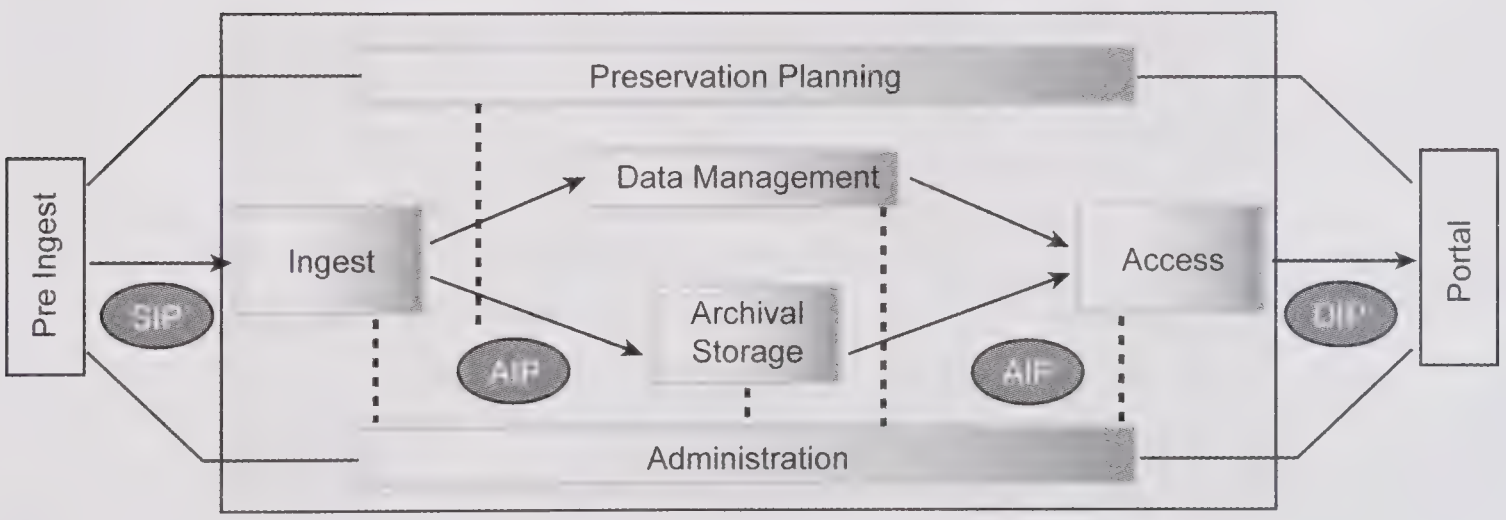


Fig. 2 – High-level overview of the OAIS components (grey box) with BHL-Europe Pre Ingest and Portal. The OAIS reference model differentiates between three kinds of information objects. The SIP, Submission Information Package, is being sent in by the data producers (content providers), the AIP, Archive Information Package, is preserved in the Archival Storage, and the DIP, Dissemination Information Package, is provided to the consumers.

Since June 2010 BHL-Europe content is made accessible for a wider public via Europeana (http://www.europeana.eu), the virtual European library. More than 80,000 books are currently accessible in Europeana and this number will increase continuously while BHL-Europe is harvesting digital literature from its content providers.

3 BHL-Europe for taxonomists

3.1 Taxonomists as portal users

A major goal of BHL-Europe is providing and facilitating open access to taxonomic literature for a number of target users (scientists, hobby scientists, students, teachers, environmental and conservation agencies, etc.). This will result in a multilingual access point for search and retrieval of biodiversity content through a robust biodiversity community portal with an open and distributed architecture and specific functionalities, as described above. Integrated web tools like taxonomic intelligence (TI) will facilitate search for taxon-specific biodiversity information and thus improve efficiency of research in biology and access to information for a wider public. The key components for achieving this goal are excellent portal tools and the improvement of optical character recognition (OCR). OCR processing of scanned pages is the base for extracting specific terms and taxon names out of the pages. However, this is still a major challenge as automatic OCR accuracy is still not sufficient for our purpose, especially for historical literature and texts in multiple languages.

BHL-Europe aims at a deep level of language integration in the indexing system (multilingual indexing). For taxonomists key data of searchable metadata are names of organisms, biological groups and provenance. However, names of taxa and groups of organisms are inconsistent, represent different and changing scientific concepts or are vernacular names. Thus it is a major task for BHL-Europe to improve existing taxon-recognition tools (TI) for an advanced search of taxonomic information.

3.2 Taxonomists as partners

BHL-Europe is building the portal and all associated services for the users to meet their needs and requirements. BHL-Europe has to understand and evaluate the requirements of the users and how they are going to use the results of the project. Therefore, a very close cooperation between the users and the project is essential to make the project a success.

BHL-Europe is targeting a large number of different users ranging from libraries over different types of scientists to the general public. A number of instruments are currently used or will be used for the user interaction to prioritise the technical and collection development plan:

(1) Web analytics will be used to quantify the use of the portal (visits, unique visitors, page views, referring sites, country coverage).

(2) Users are encouraged to drop feedback messages either using the BHL online discussion forum or using the online contact form. BHL also has an issue tracking system (Gemini) in place to collect user feedback.
(3) Face-to-face and virtual interactions between the BHL-Europe members are helpful to get important input, as the project includes a number of key users from different user groups (e.g. libraries, taxonomists). These users from within the project work together in Use Case Workgroups. Their major task is the development of use cases for the portal prototype and testing of the portal functionalities.
(4) Suggestions of actors and users of large international projects like EOL or Europeana that are setting priorities based on their experiences are taken into account.
(5) BHL-Europe considers developments in biodiversity informatics and networked scientific communication like TDWG developments or PLoS Biodiversity Hubs.
(6) Specific user evaluations will be carried out twice during the project. The results of the first online user survey analysing the demand and service elements of the project will be publicly available soon and will be fed into the BHL-Europe IT development plan.
(7) BHL-Europe offers training opportunities on how to use the portal and its functionalities as well as other BHL-Europe products, e.g. the GRIB, and will ask for feedback on possible improvements. A first workshop will be held during the BioSystematics 2011 in Berlin (http://www.biosyst-berlin-2011.de/).
(8) BHL-Europe is also present with talks and posters in numerous scientific conferences to personally discuss with the scientists and to attract new users.

All information collected from the user's side this way will form the basis for developing a comprehensive set of use cases for the BHL-Europe portal and leads to further improvements of the system infrastructure.

In the past, individual scientist or the scientific community could not influence the choice of biodiversity literature for major scanning activities. BHL-Europe is implementing a mechanism that will enable users/ scientists placing a scan request for a specific volume using the GRIB infrastructure. This will allow libraries/ content provider to set up a priority list for their scanning activities and making highly demanded literature available first. As an intermediate solution, BHL has implemented a scanning request form in their feedback system.

Another goal of BHL-Europe is to seek for new partners (content provider, right holders) that can potentially contribute open access digitised biodiversity literature to the overall BHL repository. Therefore, support from the scientific community is highly welcome in naming additional repository of digital content and bibliographic data or in discussing with rights holder to make their content freely available.

4 Conclusion

The Chinese Academy of Science and the Atlas of Living Australia have been joining BHL already and negotiations with organisations in other countries are underway to further extend the BHL network. All these projects will work

together sharing content, protocols, services and digital preservation practices and promote the idea of a Global Biodiversity Heritage Library.

Acknowledgement

BHL-Europe is co-funded by the Community Programme *e*Content*plus*, which is gratefully acknowledged.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 49-51.
ISBN 978-88-8303-295-0. EUT, 2010.

A Pan-European Species-directories Infrastructure (PESI)

Yde de Jong

Abstract — This paper introduces the rationale and aims of the Europe-wide biodiversity informatics PESI [1] project. PESI defines and coordinates strategies to enhance the quality and reliability of European biodiversity information by integrating the infrastructural components of four major community networks on taxonomic indexing, namely those of marine life, terrestrial plants, fungi and animals, into a joint work programme. This will include functional knowledge networks of both taxonomic experts and regional focal points, which will collaborate on the establishment of standardised and authoritative taxonomic (meta-) data. In addition PESI will coordinate the integration and synchronisation of the European taxonomic information systems into a joint e-infrastructure and the creation of a common user-interface disseminating the pan-European checklists and associated user-services results.

Index Terms —biodiversity, infrastructure, taxonomy, nomenclature, standards, Europe, checklists, bio-informatics, cybertaxonomy.

1 Introduction

The correct use of names and their relationships is essential for biodiversity management; therefore the availability of taxonomically validated, standardised nomenclatures is fundamental for biological e-infrastructures. PESI is the next step in integrating and securing taxonomically authoritative species name registers, serving to underpin the management of biodiversity in Europe.

PESI is a joint initiative of two Networks of Excellence: EDIT (European Distributed Institute of Taxonomy) [2] and MarBEF (Marine Biodiversity and Ecosystem Functioning) [3], funded by the European Commission under the Seventh Framework Capacities Work Programme - Research Infrastructures - and is led by the University of Amsterdam. It was started in May 2008 and will last three years, involving 40 partner organisations from 26 countries and several non-contracted associated partners.

The author is with the Zoological Museum Amsterdam, Faculty of Science - University of Amsterdam, Amsterdam, P.O. Box 94766, NL-1090 GT Amsterdam, The Netherlands. E-mail: yjong@science.uva.nl.

2 Integrating Infrastructures

2.1 Rationale

PESI defines and coordinates strategies to integrate the infrastructural components of four major community networks on taxonomic indexing and their respective knowledge (social and technical) infrastructures; those of marine life, terrestrial plants, fungi and animals, into a joint work programme. These include the three main all-taxon registers in Europe, namely the *European Register of Marine Species* [4], *Fauna Europaea* [5], and *Euro+Med PlantBase* [6] in coordination with EU-based nomenclators, i.e. *Index Fungorum* [7], IPNI [8], and *AlgaeBase* [9], plus the network of EU-based Global Species Databases (GSDs).

2.2 Coordination and integration of European expert networks

The integration of the social expertise networks will result in functional knowledge systems of taxonomic experts and regional focal points, which will collaborate on the establishment of standardised and authoritative taxonomic (meta-)data and the development of approaches for long-term data government. The sustainability of these taxonomic expert networks is considered as the most threatening issue to PESI's success, since Europe is experiencing a decline in its number of professional taxonomists.

PESI is addressing this concern by advancing the abilities of the *Society for the Management of Electronic Biodiversity Data* (SMEBD) [10], by collaborating with the *European Distributes Institute on Taxonomy* (EDIT) project and the *Consortium of European Taxonomic Facilities* (CETAF) [11], as well as reaching out to non-professional taxonomists and taxonomic societies in a hope to revive this vital science.

2.3 Coordination and integration of information e-infrastructures

The technical integration of these checklists into a joint 'European Taxonomic Backbone' relies on the *Common Data Model* (CDM) [12], ensuring the conceptual mapping of taxonomic databases. This is hosted in the CDM store as a denormalised relational database management system (the so-called 'PESI data warehouse'). The CDM represents a component of EDIT's Cybertaxonomy Platform [13].

PESI is also involved in supporting international efforts on the development of the 'Global Names Architecture' (GNA) [14] by building a common intelligent name-matching device in consultation with principal initiatives like GBIF [15] and LifeWatch [16]. This provides a unified cross-reference system to all stakeholders optimising their taxonomic meta-data service functioning.

2.4 Integrated e-Services for users and dissemination

PESI will build an interactive, multilingual web portal [17] to carry out the dissemination of the developed species names service and to support the use of the pan-European species data in the e-science domain. This will include relevant supplementary data, like occurrence details by applying dynamic links to pertinent e-data services.

Acknowledgement

The authors wish to thank all PESI partners, especially PESI work-package leaders and managers, for their contributions.

References

[1] PESI (http://www.eu-nomen.eu/pesi)
[2] EDIT (http://www.e-taxonomy.eu)
[3] MarBEF (http://www.marbef.org)
[4] ERMS (http://www.marbef.org/data/erms.php)
[5] Fauna Europaea (http://www.faunaeur.org)
[6] Euro+Med PlantBase (http://www.emplantbase.org/home.html)
[7] Index Fungorum (http://www.indexfungorum.org) also (http://pesi.indexfungorum.org)
[8] IPNI (http://www.ipni.org)
[9] AlgaeBase (http://www.algaebase.org)
[10] SMEBD (http://www.smebd.eu)
[11] CETAF (http://www.cetaf.org)
[12] CDM (http://dev.e-taxonomy.eu/trac/wiki/CommonDataModel)
[13] Cybertaxonomy Platform (http://wp5.e-taxonomy.eu)
[14] GNA (http://www.gbif.org/informatics/name-services/global-names-architecture)
[15] GBIF (http://www.gbif.org)
[16] LifeWatch (http://www.lifewatch.eu)
[17] PESI portal (http://www.eu-nomen.eu/porta)

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 53.
ISBN 978-88-8303-295-0. EUT, 2010.

ViBRANT—Virtual Biodiversity Research and Access Network for Taxonomy

Dave Roberts, Vince Smith

Abstract — Biodiversity science brings information science and technologies to bear on the data and information generated by the study of organisms, their genes, and their interactions. ViBRANT will help focus the collective output of biodiversity science, making it more transparent, accountable, and accessible. Mobilising these data will address global environmental challenges, contribute to sustainable development, and promote the conservation of biological diversity. Through a platform of web based informatics tools and services we have built a successful data-publishing framework (Scratchpads) that allows distributed groups of scientists to create their own virtual research communities supporting biodiversity science. The infrastructure is highly user-oriented, focusing on the needs of research networks through a flexible and scalable system architecture, offering adaptable user interfaces for the development of various services. In just 28 months the Scratchpads have been adopted by over 120 communities in more than 60 countries, embracing over 1,500 users. ViBRANT will distribute the management, hardware infrastructure and software development of this system and connect with the broader landscape of biodiversity initiatives including PESI, Biodiversity Heritage Library (Europe), GBIF and EoL. The system will also inform the design of the LifeWatch Service Centre and is aligned with the ELIXIR and EMBRC objectives, all part of the ESFRI roadmap. ViBRANT will extend the userbase, reaching out to new multidisciplinary communities including citizen scientists by offering an enhanced suite of services and functionality.

The authors are with the Natural History Museum,Cromwell Road, London, SW7 5BD, UK. E-mail: vince@vsmith.info, dmr@nomencurator.org.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 55-57.
ISBN 978-88-8303-295-0. EUT, 2010.

Identifications in BioPortals™

Wouter Addink, Edwin van Spronsen, Peter H. Schalk

Abstract — BioPortals are a 'Google-like' webportal solution tailored for national or thematic biological diversity information needs. This solution allows for an efficient route to retrieve information from heterogeneous biological information sources after identification with integrated identification systems. BioPortals can be used in combination with mobile devices and provide options to share biological observation data after identification in the field.

Index Terms — BioPortals, species information, identification, webportal, mobile platforms.

1 Introduction

Digital information on biological diversity (i.e. on identification, species, taxonomy, ecology, genetics, conservation, legislation) is often compiled for a specific purpose and stored in custom-made, geographically distributed software using different formats. Mining this information can therefore be time consuming, and recombination of data can be cumbersome because of incompatibilities. Furthermore, a proper identification of the species is required in order to be able to retrieve the right biological information.

International initiatives, such as the Global Biodiversity Information Facility (GBIF), the Encyclopedia of Life (EoL), and the Consortium for the Barcode of Life (CBOL), are paving the way to make data in a specific domain globally accessible. However, much of the demand for information is on a national or thematic level, driven by defined groups of users with specific questions or problems. These call for a custom-made answer to their information needs.

ETI developed a 'Google-like' webportal solution that provides a single access point to a large array of heterogeneous biodiversity information sources, combining them with identification keys. The so-called *BioPortal* can be customized to specific information needs and user levels: from scientists to conservationists, from governments to schools.

2 Identification Keys

Keys are a method to identify by asking a series of questions to the user.

W. Addink is Head of Informatics, ETI BioInformatics, Amsterdam. wouter@eti.uva.nl.
E. van Spronsen is Head of Information, ETI BioInformatics, Amsterdam. edwin@eti.uva.nl.
P. Schalk is Director of ETI BioInformatics, Amsterdam. peter@eti.uva.nl.

Each answer excludes some names, until the most likely name for the species remains. This name can then be compared with the description of the species to confirm the identification. This method is more efficient than going through a series of species descriptions until a description has been found that fully matches the observed species.

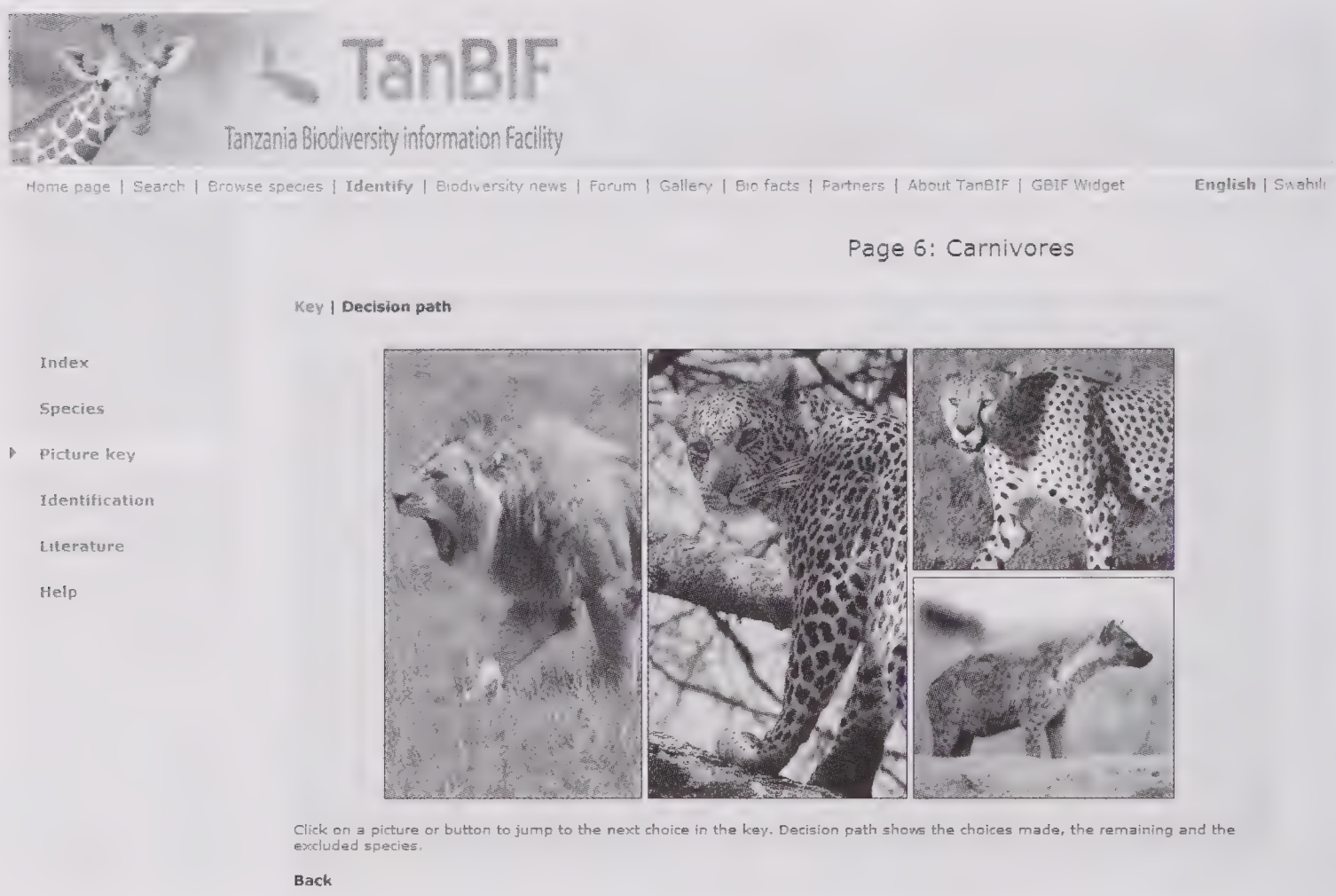

TanBIF
Tanzania Biodiversity information Facility
Home page | Search | Browse species | Identify | Biodiversity news | Forum | Gallery | Bio facts | Partners | About TanBIF | GBIF Widget
English | Swahili
Page 6: Carnivores
Key | Decision path
Index
Species
Picture key
Identification
Literature
Help
Click on a picture or button to jump to the next choice in the key. Decision path shows the choices made, the remaining and the excluded species.
Back

Fig. 1 – A picture key in the Tanzanian national biodiversity portal, built with the BioPortal toolkit.

Traditional keys published on paper are usually dichotomous, text-based keys, where the user is given a sequence of choices, each choice leading to a new choice and excluding some names until only one name remains. The BioPortal toolkit supports such traditional keys, but also other types of keys that allow for more interaction and easier identification. In such keys the user can select only the characters that are visible, for example only leaf characters to identify a plant that is not flowering, or by using suggestions for the characters that are the most separating ones. Keys that use illustrations or sounds instead of text, or combinations of these are also supported.

3 From Identification to Information

The BioPortal toolkit includes the *Linnaeus II* Species Bank compiler that allows the creation of targeted (e.g., national) species information systems with identification tools for the Internet (e.g. www.soortenbank.nl), accessible with web browsers on computers or on mobile devices such as iPad, PDA or iPhone. The Linnaeus II compiler is compatible with the systems and exchange formats developed in the *KeyToNature* project. Identification keys that have been made available in *KeyToNature* can therefore be used in BioPortals. This allows users

of BioPortals to go directly from an identification made with a *KeyToNature* identification key to information about the identified species from a range of online information sources.

4 Sharing Information after Identification

With mobile identification tools like the Netherlands e-Flora application for iPhone, identifications can be made on spot. When a species is identified, the species observation (identified species name, photo, date, long-lat location, observer) can be uploaded directly to a central server and, after vetting, displayed in a BioPortal. This allows BioPortal users to see the identified observations in the portal together with related information from other sources.

5 *BioPortal* Design and Implementation

The *BioPortal* toolkit has a modular design that can be tuned to user needs, including customizing the web-interface and providing support for multiple languages. The toolkit combines content management system (CMS) functionality, such as news items and static pages, event tracking, and forum modules, with functions to access the scientific core data. These are biological collections, observations, ecological relations, molecular data, and image libraries compiled on a national or local level, combined with external sources such as GBIF or EMBnet (European Molecular Biology Network). Sophisticated technology allows for simultaneous server-side asynchronous searches in several distributed data sources on the web. The Catalogue of Life is used as a validated taxonomic index to match searches to connected data sources.

ETI's *BioPortal* toolkit was used to build the Netherlands' national GBIF portal (NLBIF), http://www.nlbif.nl and the Tanzanian national biodiversity portal (TanBIF: http://www.tanbif.or.tz). It is also used to build the global pollinators portal.

Acknowledgement

The authors wish to thank all the persons involved in *KeyToNature* throughout Europe. Their efforts and input gave use new ideas and energy to develop them. This paper was produced in the framework of the the project *KeyToNature* (www.keytonature.eu, ECP-2006-EDU-410019), funded in the *e*Content*plus* Programme.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 59-64.
ISBN 978-88-8303-295-0. EUT, 2010.

Types of identification keys

Gregor Hagedorn, Gerhard Rambold, Stefano Martellos

Abstract — A number of terms related to identification tools are introduced and the advantages of selected types of identification keys are compared.

Index Terms — Identification tools, single-access key, dichotomous key, polytomous key, lead, couplet, free-access key, multi-access key, matrix key, multi-entry key.

1 Introduction

The generalization of individuals (things, events, etc.) into classes is essential to transfer knowledge across individual incidents. When learning a language, we learn the defining features of classes like "table", "chair", "shrub", etc. Similarly, biology defines formal classes for living things (called "taxa") together with class names ("taxon names") and defining descriptions. The assignment of an unknown object to a taxon is called "identification" or "determination". To non-biologists this may be confusing, the term "identification" being more commonly associated with the naming of individuals (as in "ID card" or "record identifier").

The number of taxa in biology is very large. For example, currently about 900 000 insect taxa alone are recognized. Compared to the average vocabulary of an educated English native speaker of roughly 25 000 words, it is clear that teaching the vast "taxon vocabulary" to biology students was always problematic. Although comparing a collected specimen sequentially with published descriptions or representative specimens is an essential identification method, any "linear search" method comparing one specimen after another soon becomes impractical.

Biologists have therefore developed various forms of "identification keys" to "unlock" the knowledge that would otherwise remain inaccessible. These are essentially "divide and conquer" search algorithms that reduce the result set recursively until the remainder is small enough to be solved by direct comparison. The fastest algorithms are those that provide a division into equally sized partitions (leading to search algorithms that scale logarithmically with

G. Hagedorn is with the Julius Kühn-Institute, Federal Research Centre for Cultivated Plants, Institute for Epidemiology and Pathogen Diagnostics, Königin-Luise-Straße 19, D-14195 Berlin, E-mail: gregor.hagedorn@jki.bund.de. – G. Rambold is with the Department of Mycology, University of Bayreuth, D-95440 Bayreuth – S. Martellos is with the Dept. of Life Sciences, Univ. of Trieste, I-34127, Trieste, Italy. E-mail:martelst@units.it.

the number of taxa). Biological keys don't always provide this, because other factors (character observation reliability, convenience, cost, etc) conflict with the desire to provide fastest progress. The authors of biological identification keys, however, typically realize that evenly splitting choices are desirable.

2 Single-access, multi- or free-access, and multi-entry keys

The most traditional biological identification keys are the easily printable dichotomous (every choice has two alternatives) or polytomous (two or more options at each choice) forms. The structure of these keys typically consists of a series of alternative statements, called "leads". All leads that need to be evaluated for a single decision form a "couplet".

Clearly, dichotomous keys are a special case of polytomous keys. For simple choices involving a single characteristic (a "character" or "feature"), such as "wings: 1. present 2. absent" or "flower colour: 1. red 2. blue 3. pink 4. yellow", the number of options matters little. However, to achieve reliable identification in the face of natural variability and continuous variation, it may be desirable to use complex Boolean statements involving multiple characters. For example, "leaves hairy and flowers red" versus "leaves glabrous or flowers not red" may be used in a case where the alternative may include glabrous plants with pinkish-red flowers, or plants with mixed hairiness but other flower colours. Although it is theoretically possible to construct a polytomous key with Boolean lead statements, practice has shown that the result is often akin to a logical riddle. Many editors therefore recommend or require the use of dichotomous keys.

However, a key may be a mixture of simple polytomous and complex dichotomous choices. The generalizing term "single-access key" [1] is therefore used in the present paper to include both dichotomous and polytomous keys. The equivalent term in computer science is "decision tree".

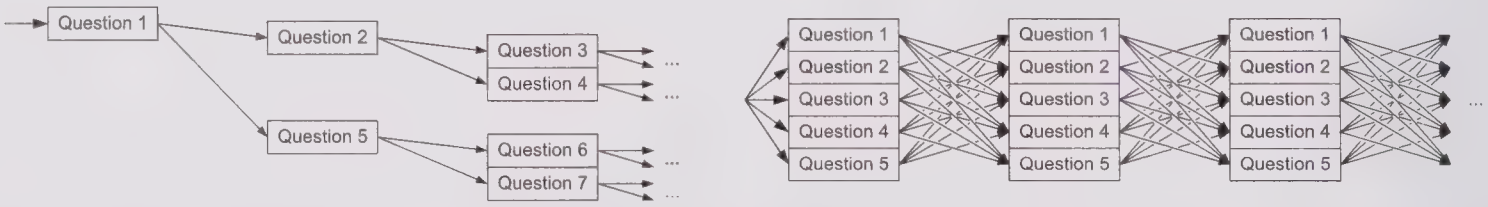


Fig. 1 – User interaction steps in a single-access key (left, the sequence of steps follow the data structure) and a free-access key (right, the sequence is determined by the user). From [2].

An alternative to a single-access key is the *free-access key* (also known as multi-access key, matrix key, or, incorrectly, "synoptic key"[1]). Whereas in a single-access key a fixed sequence of choices (decisions) is defined by the author (providing a single path to each result), in a free-access key the sequence of choices is up to the user. In every step, the user can select from the list of characters offered, and choose a matching state or value. Thus a free-access key is the set of all possible single-access keys that arise by permutating the order of characters (Fig. 1). Although printable free-access keys exist, they

[1] The term „synoptic key" has traditionally been used for single-access keys that reflect the taxonomic hierarchy. Its use for multi-access keys (especially printable ones) should be avoided.

are most suitable for computer-aided identification tools, and have a long development history [3]. Examples are DELTA-IntKey [4], Lucid [5], NaviKey [6], Xper^2 [7]. The Flash-based IBIS-ID [8] was newly developed in *KeyToNature*.

In a free-access key, the choice of characters is repeated at every step. A related form, the multi-entry key, allows free choice of characters in a first step (a "multi-character-query-form"), followed either by a field-guide-like listing of remaining taxa, or by a dynamically generated (filtered) single-access key (as in the FRIDA/Dryades keys [9]).

	Single-Access	**Free-Access**	**Multi-Entry**
Information reduction	High	None (complete information is optimal)	Variable (none if all characters are available in the initial step)
Average identification speed	Depends on the creators of the key	Depends on user's background knowledge; may exceed average	Variable between single and multi-access
Complex statements (and, or, etc.)	Yes (not recommended for polytomous keys)	No	"No" in entry-form, "yes" in following single-access key
Question-answer style	Possible for simple statements	(Implicit in character state or value choice)	(Implicit in character state or value choice)
Difficulty of choosing next decision	None	Often high for beginners	Variable; depends on completeness of initial entry form
Skipping unanswerable choices	Difficult; all alternative paths must be followed to the end	Easy	Easy in entry-form, difficult in an optional single-access part
Resources required for construction	Low for first draft. Good keys require high expertise	High investment until first version can be tested	High for data matrix, but the size of the matrix is variable

Tab. 1 – Comparison of some identification methods. The comparison is aimed at manually created single-access keys (those generated from a data matrix are not considered here).

An evenly splitting single-access key requires fewer decisions from beginners than a multi-access key. The latter may in fact generate faster progress in terms of "steps", but requires additional decisions as to which character to use next. Even if the character list is ordered by character suitability and fastest progress, beginners will be tempted to use execute their free character choice. This is problematic if many characters are not yet understood. In contrast, the fixed

identification path of a single-access key also fixes which terms and concepts must be learned first. A disadvantage of single-access keys is that identification may be impossible if a choice cannot be decided at all. This may occur because a character cannot be observed (e. g., a developmental stage is not present in the specimen), or because the options are not communicated clearly enough. The resulting frustration can be high, especially for beginners.

Both free-access and multi-entry keys truly excel in their performance when used by experts. For these, character selection is intuitive and fast. By choosing characters, for which a rare state is present in the specimen, identification progress can then be on an order of magnitude faster than using a single-access key. This is already possible with moderate experience, since states that were never observed by a user before are, by definition, rare. Tab. 1 gives an overview of some differentiating features.

From an author's perspective, matrix-based keys require a high initial investment to research and fill a large character × taxon matrix. In contrast, single-access keys require less formal investment. Due to the inherent information reduction (most characters apply only to a relatively small subset of taxa), a reviewable key is faster to produce and proof-reading is less time-consuming than the creation of an equivalent data matrix containing all characters for the same group of taxa. However, a successful single-access key depends strongly on the expertise of the author to chose characters that are convenient, cost-effective, reliable across all taxa in the subtree, and available throughout a large period of the developmental cycle of the organism. Single-access keys may therefore require several cycles of testing until initially overlooked problems have been fixed; their production can be akin to the "debugging" of software code.

Furthermore, the creation of matrix-based keys generally requires learning a special-purpose application like DELTA or Lucid, whereas single-access keys may be created in a text-processing application. Therefore, although newly created single-access keys may occasionally be problematic to use, they offer considerable benefits to both producers and consumers.

Single-access keys, until recently, have been developed only rarely as computer-aided, interactive tools. Noteworthy developments in this direction are the commercial Lucid Phoenix application [10], the FRIDA/Dryades software [9], [11], the *KeyToNature* Open Key Editor" [12], and the open source WikiKeys and jKey [13] application on biowikifarm [14].

3 Structural variants of single-access keys

Two additional structural variants of single access keys are relevant when building information models [2]: 1. Couplets may consist of a question, with the leads providing contrasting answers. This question-answer-style is often appealing to beginners. However, complex statements (involving more than one character and Boolean expressions) are not possible. Whereas a mixture of simple polytomous and complex dichotomous couplets in a single key is quite intelligible, a mixture of simple question-answer-style with question-less complex dichotomous couplets is not. 2. The desire for fast identification progress using

convenient characters often conflicts with character variability in a subset of organisms. As long as the character is reliable for the majority of taxa, a frequent solution to this problem is to key out taxa with variable character expression multiple times. This may affect only the terminal taxa, or entire branches of the keys. Whereas the first case will often simply be handled by true duplication, multiple references to entire branches of a decision tree turn a "tree" structure into a directed (and generally acyclic) graph (DAG) and requires careful attention when modelling information models or software. In biology a DAG is sometimes called a "reticulated" identification key.

Linked Key Style (also called "parallel", "juxtaposition" or "bracketed" style):

1. Ovule solitary, basal **2**
– Ovules numerous, axile or free-central **3**
2. Perianth green, membranous or absent; filaments free .....**Chenopodiaceae**
– Perianth translucent and papery; filaments often united below .....**Amaranthaceae**
3. Placentation axile; leaves alternate .....**Saxifragaceae**
– Placentation basal or free-central; leaves usually opposite **4**

Nested Key Style (also called "yoked" or "indented" style):

1. Ovule solitary, basal
  2. Perianth green, membranous or absent; filaments free .....**Chenopodiaceae**
  2. Perianth translucent and papery; filaments often united below .....**Amaranthaceae**
1. Ovules numerous, axile or free-central
  3. Placentation axile; leaves alternate .....**Saxifragaceae**
  3. Placentation basal or free-central; leaves usually opposite

Fig. 2 – Examples of the linked and nested styles of branching keys in lead style; see [2] for derivation.

4 Presentational variants of single-access keys

The dominant presentation styles of single-access keys are shown in Fig. 2. In "linked" keys the connection between couplets is achieved by a linking reference (at the right side) to a couplet ID (left). In nested keys direct nesting of couplets replaces the explicit linking. Nested keys are more commonly known as "indented", but unfortunately this refers to an accidental (albeit frequent) rather than essential quality [3]. Nested keys may be printed without indentation to preserve space (relying solely on corresponding lead symbols) and linked keys may be indented to enhance the visibility of the couplet structure. Further presentational ("solid keys", "graphical style") and semantic ("artificial" or "diagnostic" versus "natural", "synoptic", or "phylogenetic") variants exist; see [2].

5 Summary

The order of couplets (choices) in an identification tool may be defined by the creator (single-access key), or may be freely selectable by the user (free-access key). A multi-entry key is an intermediate form that may combine advantages of both forms if only a small character subset is included in the multi-entry phase.

Structural criteria for single-access keys are: a) whether the leads in a couplet are limited to two (dichotomous) or not (polytomous key); b) whether couplets are limited to a single character or combinations of multiple characters, involving Boolean operators such as 'and', 'or', or 'not', are supported; c) whether taxa may be keyed out in multiple places, and whether redirections into entire sections (or "branches") of the key are supported ("reticulated key"); and d) whether leads in couplets are complete statements or split into a question with the couplet and leads providing the answers. Certain presentational forms (nested key versus linked keys) are not structurally relevant.

Acknowledgement

This work was supported by the *KeyToNature* Project, ECP-2006-EDU-410019, in the *e*Content*plus* Programme. We thank P. L. Nimis and B. Press for review and important suggestions.

References

[1] J. Winston, *Describing Species*. Columbia University Press, 1999.

[2] G. Hagedorn, *Structuring Descriptive Data of Organisms — Requirement Analysis and Information Models*. Ph. D. Thesis, Universität Bayreuth, 2007.

[3] R. J. Pankhurst, *Practical Taxonomic Computing*, 1991.

[4] DELTA – DEscription Language for TAxonomy http://delta-intkey.com/, 2010-07.

[5] Lucidcentral.org http://www.lucidcentral.com, 2010-07.

[6] D. Neubacher and G. Rambold, NaviKey – a Java applet and application for accessing descriptive data coded in DELTA format, 2005 (onwards). http://www.navikey.net, 2010-07.

[7] V. Ung, G. Dubus, R. Zaragüeta-Bagils and R. Vignes Lebbe, "Xper^2: introducing e-taxonomy". *Bioinformatics*, 26 (5): 703-704, 2010; doi: 10.1093/ bioinformatics/btp715; see also http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2.

[8] M. Giurgiu, G. Hagedorn and A. Homodi, "IBIS-ID, an Adobe FLEX based identification tool for SDD-encoded multi-access keys". *Proc. of TDWG 2009*, 9-13 Nov. 2009, Montpellier, p. 90, 2009.

[9] S. Martellos and P. L. Nimis, "*KeyToNature*: Teaching and Learning Biodiversity. Dryades, the Italian Experience". In: M. Muñoz, I. Jelìnek, F. Ferreira (eds.), *Proceedings of the IASK International Conference Teaching and Learning 2008*, pp. 863-868, 2008.

[10] Lucid Phoenix (http://www.lucidcentral.org/LinkClick.aspx?link=152), (2010-07).

[11] S. Martellos, "Multi-authored interactive identification keys: The FRIDA (FRiendly IDentificAtion) package", *Taxon,* vol. 59 (3), pp. 922-929, 2010.

[12] S. Martellos, E. v. Spronsen, D. Seijts, N. Torrescasana Aloy, P. Schalk and P. L. Nimis, "User-generated content in the digital identification of organisms: the *KeyToNature* approach", *Int. J. Information and Operations Management Education*, vol. 3, 3, pp. 272-83, 2010.

[13] S. Opitz, G. Hagedorn, "The jKey wiki key player and builder". *Proc. of TDWG 2009*, 9-13 Nov. 2009, Montpellier, 2009.

[14] G. Hagedorn, G. Weber, A. Plank, M. Giurgiu, A. Homodi, C. Veja, G. Schmidt, P. Mihnev, M. Roujinov, D. Triebel, R. A. Morris, B. Zelazny, E. van Spronsen, P. Schalk, C. Kittl, R. Brandner, S. Martellos and P. L. Nimis, "An online authoring and publishing platform for field guides and identification tools". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 13-18, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 65-70.
ISBN 978-88-8303-295-0. EUT, 2010.

Learning, Identifying, Sharing

Philippe A. Martin, Noël Conruyt, David Grosser

Abstract — This article argues that a cooperatively-built, well-organized, shared knowledge base is a new – and, from certain viewpoints, optimal – kind of support (refining and integrating other kinds of supports) for three complementary tasks: learning about living entities (and how to identify them), supporting their identification, and sharing knowledge about them. This article gives the ideas behind our prototype, and argues that knowledge providers can be not solely specialists, but also amateurs. In essence, for these three tasks, it argues for the (re-)use of much more semantically organized and interconnected versions of semantic wikis or scratchpads.

Index Terms — identifying, knowledge sharing, learning, ontologies, semantic wikis.

1 Introduction

Current supports for learning about – and identifying – living entities, e.g., the supports listed by the *KeyToNature* project (www.keytonature.eu), are mostly static files (texts, images, ...) and tools based on a formal[1] knowledge base (KB). Few tools allow their users to contribute annotations or other information to their formal or informal KB, let alone use them for i) helping identification or learning, and ii) publishing them in a way usable by other tools. Scratchpads [1] and, more generally, semantic wikis[2], allow the cooperative edition and semantic linking of information by any web user, but not in an organized or formal enough way to be re-used by an identification tool or a problem-solving tool, nor to permit the automatic detection of partially redundant/inconsistent information within or between wikis. This automatic detection is essential to permit the semi-automatic and cooperative organization of knowledge into a ***unique semantic network*** and thus permit i) ***scalable information retrieval, comparison, sharing and exploitation***, and hence ii) an ***easier understanding or learning (by amateurs or specialists)*** of the stored information and viewpoints of their authors. Section 2 quickly compares the various current kinds of supports for the learning and sharing of information

The authors are with the IREMIA laboratory, University of La Réunion. E-mail: (Philippe.Martin, Noel.Conruyt, David.Grosser)@univ-reunion.fr.

[1] In this article, "formal" means *machine processable and logic-based*, while "semantic" means *formal and organized by semantic relations, e.g., "subtype of", "physical_part of", "agent of" and "duration of".*

[2] Semantic wikis are collaboratively-built documents with some parts indexed by semantic categories or interconnected by semantic relations. See http://semwiki.org *for more details.*

about living entities and hence for helping their identification.

Section 3 introduces elements required to support an approach leading to a ***global KB*** composed of collaboratively-built KBs that have no *implicit*[3] "automatically detectable partial redundancies or inconsistencies" neither within nor between the KBs. As suggested in Section 2, such a global KB – and hence this approach (which is complementary to the other approaches) – is the most useful one from a knowledge-sharing, retrieval and learning viewpoint, but its disadvantages are that i) it requires the users to learn how to read a textual or graphic notation for representing or interconnecting knowledge, and ii) for each domain that has not yet been well represented in the shared KB, the first knowledge providers have a lot of work to do for organizing the information resulting from the use of other approaches. However, this can be done incrementally, whenever the benefits finally becomes clearer than the costs. The elements of this approach are fully or partially implemented in our knowledge server WebKB-2 [2] (webkb.org).

2 Quick Comparison of Approaches

The ***smaller*** the sources of information used for knowledge sharing – i.e. the less objects of information (e.g., statements or images) these resources contain – and the ***less contextual*** (hence more explicit, precise and formal) these objects are, the easier it is to automatically index these resources precisely, to filter out the redundancies and to relate these resources via semantic relations, e.g., to organize them into a specialization hierarchy[4]. Then, the ***easier*** it is to ***retrieve*** these resources (by querying or browsing)[5], ***compare*** them (hence, understand and memorize them), ***combine*** them and, more generally, ***exploit*** them for various purposes, e.g., guiding ***identification***. As illustrated in the following paragraphs, these rather obvious ideas are generally well accepted, but their ***ultimate conclusion*** is socially and technically difficult to bring about and hence not directly studied. The conclusion is: there should *ideally* be one and only ***one global semantic network*** (i.e., each index or symbolic resource should contain only one statement or one formal term; in other words, there should be no difference between should symbolic data and meta-data) and, in this network, all manually or automatically detected partial redundancies or inconsistencies are made explicit via semantic relations. In this article, such a global semantic network is called a ***global cbwoKB*** (cooperatively-built well-organized KB).

The ***Learning object*** (LO) related community and standards (e.g., IEEE LTSC)

---

[3] In this article, implicit means "not made explicit via a semantic relation".

[4] Related small *individual* statements can often be organized into a specialization hierarchy or an inclusion hierarchy but *sets* of related statements rarely can (the bigger the sets, the less likely).

[5] For example, if the query is of the kind "what are the resources/tools/methods to do ...", the answer can be a part/subtask/specialization hierarchy (with associated argumentation structures). Such semantically structured answers allow a user to find and compare all relevant objects instead of getting a long list of partially redundant objects or files where original/precise ones are hidden among/behind objects that are more general, more mainstream or from big organizations.

advocate the use of small non-contextual LOs but still only considers the use of static informal documents indexed by keywords. Semantic LO repositories [3] use formal terms or statements for indices. This is also the approach used by STERNA [4].

As highlighted in [5] and [6], the ***Semantic Web*** (SW) community currently essentially focuses on inference mechanisms, KB editors, semantic wikis, social networks, workflow-based cooperation, and the semi-automatic partial interconnection of the content of (semi-)independently created KBs or formal files. Tools created by this community do not *directly* support the creation of a cbwoKB (global or local) and, in a sense, they participate to the problems they are trying to solve since their outputs create new files that are partially redundant or inconsistent with their input files and without semantic relations to make this explicit. The current focus of the SW community is to work with approaches hiding the knowledge representations from the users as much as possible. The problem is then that the semantic network cannot be completed in a meaningful way by the users (only low quality knowledge can be automatically extracted and exploited) nor even browsed to find information. As an example, ***semantic wikis*** are still mainly poorly organized informal documents. Instead, in WebKB-2 the semantic network can be edited by all Web users via cooperation protocols and can be viewed in a more or less structured way via various relatively intuitive syntaxes [7]: Formalized-English, For-Links, etc. However, reading these syntaxes requires a short training and writing knowledge requires the following of some given conventions or "best practices".

Scratchpads are kinds of semantic wikis which, according to some of their documentation [8], are "independent and unconnected, allowing communities to create distinct customized sites tailored to their needs". This strongly reduces the possibilities of (semi-)automatically comparing and integrating the content of different scratchpads, and hence works against the goals of identification-related projects like ViBRANT [9] which is based on the use of scratchpads. With a cbwoKB, tailoring can be done by *each* user using filters and presentation rules.

Many identification-related projects use ***databases***, e.g., FishBase (fishbase.org) and Pl@ntNet (plantnet-project.org). They have a regular structure but a rather flat one and users cannot directly contribute to the database: annotations, new objects, new tables (classes of objects), new attributes (relations from/to objects), etc. Finally, the semantics of the objects of these databases is unknown unless their semantic relations to other objects from the Semantic Web are described in a formal file.

Except for WebKB-2, current ***KB servers/editors*** (e.g., Ontolingua, OntoWeb, Ontosaurus, Freebase, CYC and semantic wiki servers) have no shared KB editing protocols and hence either i) let every authorized user modify what other ones have entered (this discourages information entering or leads to edit wars), or ii) require all/some users to approve or not changes made in the KB, possibly via a workflow system (this is bothersome for the evaluators, may force them to make arbitrary selections, and this is a bottleneck to information sharing that often discourages information providers). To complement the generic "knowledge sharing" features of ***WebKB-2*** with identification features, its integration with ***IKBS*** [10], a KB based identification tool, has begun.

3 Underlying Ideas of Solutions for the Proposed Approach

To be a ***generic "knowledge sharing"*** support, the shared KB of WebKB-2 has been initialized via a ***loss-less merge of many ontologies*** (sets of formal terms with their associated definitions/constraints/inter-relations): top-level ones (including methodological ones such as DOLCE) and a lexical one (an extension and correction of WordNet) [11]. Knowledge ***normalization rules*** have been collected and extended; simultaneously, various complementary, ***expressive and relatively intuitive notations*** enforcing these rules have been designed [7]. Finally, knowledge sharing protocols have been designed [2]. The ***protocols for the collaborative edition of a shared cbwoKB*** have been implemented and are introduced in the second next paragraph. This is not yet the case for the ***protocols permitting to create a global cbwoKB*** composed of several cbwoKB servers. Their underlying idea is that each of these servers must i) publish its commitment to be a "nexus" for one or several formal terms, that is, to store all information directly related to these terms, and ii) point to other nexus for terms it is not the nexus of. In this way, via redirections of queries and replications of knowledge between servers, it does not matter which server a user updates or queries first, and the advantages of distribution and centralization are thus combined.

WebKB-2 has an ***expressive language model*** (1st-order logic, n-order types, meta-statements and collections) but has a ***simple data model*** since it is built on top of an object-oriented DBMS with only three tables: Term, Relation and Source. Every object of the KB is either a formal/informal term or a formal/semi-formal/informal statement (e.g., a relation between two quantified terms, and a relation on a relation in order to represent some spatial and temporal context). Every object has one or several associated sources: i) the user who created the object, ii) the original resource (e.g., a person, a language, a document) from which the user read/heard/took the object and hence *interpreted* it, and iii) other users who also believe in that object (if it is a statement). ***Lexical conflicts*** are avoided by prefixing formal terms with the identifier of their creators, e.g., wn#bird refers to the most common concept (i.e., meaning) proposed by WordNet for the word "bird".

The next sentences introduce the most important basic ideas behind the shared KB editing protocols of WebKB-2 and hence behind the ways ***semantic conflicts are avoided and the KB kept "well organized"***. A user can re-use any object (term or statement) but can only modify or remove an object that he has created. *Adding, modifying or removing a term* is done by adding, modifying or removing at least one statement (generally, one relation) that uses this term. *A new term can only be added by specializing another term.* Each object must be connected to at least another object via relations of specialization/generalization, identity and/or argumentation (and as many as possible of such relations should be used). If a user adds, modifies or removes a statement (definition or belief) and this creates a *detected conflict* (*redundancy and inconsistency) with another of his statements*, the action is *rejected*. If *adding, modifying or removing a (definition of) a term introduces a conflict with statements of other users,* this conflict highlights an over-interpretation of the term by these other users and

this is automatically *solved by "cloning" the term,* i.e., creating a slightly more general copy of this term for these other users to repair the over-interpretation. If *adding, modifying or removing a belief introduces a detected potential conflict* (partial/total inconsistency or redundancy) involving beliefs created by other creators, it is *rejected*. However, a user may still represent his belief (say, b1) – and thus "loss-less correct" another user's belief that he does not believe in (say, b2) – by connecting b1 to b2 via a corrective relation. E.g., here is a Formalized-English statement by u2 which corrects a statement made earlier by u1:
u2#` u1#`every bird is agent of a flight´ has for corrective_restriction u2#`most healthy flying_bird are able to be agent of a flight´.

This statement means: "according to u2, u1's belief that 'every bird flies' is false and a more precise statement is 'most healthy flying birds (the carinates) are able to fly". This way the KB is kept organized and ***then, if necessary, an inference engine can choose between such statements according to the constraints of a particular application***, e.g., it can always choose the most precise version or it can choose the one authored by someone represented as an expert in a certain domain. ***Similarly***, in the same way he creates queries, ***a user can create filters on the content, authors, ..., and popularity of statements*** in order to see *only what he wants to see* when browsing the KB. With this approach, ***every author can represent his beliefs, no selection committee is required, and knowledge integration is loss-less*** (the sources can be regenerated). This approach also ***avoids the problems related to version control or truth-maintenance***.

4 Conclusion

This article compared various knowledge sharing approaches and introduced elements necessary to support the most precision-oriented and end-user-controlled approach and the one that combines the advantages of the centralization and distribution. Thus, it is the approach that most permits to i) retrieve and compare knowledge about a living entity and hence learn about it, ii) integrate knowledge from everyone (specialists and amateurs), and iii) leads to create knowledge that directly or indirectly can be re-used by tools to guide identification. Most of these elements are implemented in WebKB-2. It will soon be used to enable Web users to extend the content of FishBase and PI@ntNet.

References

[1] V. S. Smith, S. D. Rycroft, K. T. Harman, B. Scott and D. Roberts, "Scratchpads: a data-publishing framework to build, share and manage information on the diversity of life," *BMC Bioinformatics 2009,* 10 (suppl. 14). See also http://scratchpads.eu, 2010.

[2] P. Martin, "Protocols for Governance-free Loss-less Well-organized Knowledge Sharing", *ECAI 2010 workshop on Intelligent Engineering Techniques for Knowledge Bases (I-KBET 2010)*, Lisbon, Portugal, 17 August 2010.

[3] J. S. Carrion, E. G. Gordo and S Sanchez-Alonso, "Semantic learning object repositories", *International Journal of Continuing Engineering Education and Life Long Learning*, vol. 17, 6, pp. 432-446, 2007.

[4] STERNA, "Semantic Web-based Thematic European Reference Network Application", http://www.sterna-net.eu, 2010.

[5] N. Shadbolt, T. Berners-Lee and W. Hall, "The semantic web revisited", *IEEE Intelligent Systems*, 21, vol. 3, pp. 96-101, May/June 2006.
[6] R. Palma, P. Haase, Y. Wang and R. d'Aquin, "Propagation models and strategies", *Deliverable 1.3.1 of NeOn* - Lifecycle Support for Networked Ontologies; NEON EU-IST-2005-027595, 2006.
[7] P. Martin, "Knowledge representation in CGLF, CGIF, KIF, Frame-CG and Formalized-English", *Proc. of ICCS 2002*, Springer LNAI 2393, pp. 77-91, 2002.
[8] P. Martin, "Protocols for Governance-free Loss-less Well-organized Knowledge Sharing", *Proc. ECAI 2010 workshop on Intelligent Engineering Techniques for Knowledge Bases* (I-KBET 2010), Lisbon, Portugal, 17 August 2010.
[9] ViBRANT, "Virtual Biodiversity Research and Access Network for Taxonomy", E.U. FP6 project, http://vbrant.org, 2010.
[10] N. Conruyt and D. Grosser, "Knowledge management in environmental sciences with IKBS: application to Systematics of Corals of the Mascarene Archipelago", *Selected Contributions in Data Analysis and Classification*, Springer Series: Studies in Classification, Data Analysis and Knowledge Organization, pp. 333-344, 2007.
[11] P. Martin, "Correction and extension of WordNet 1.7", *Proc. of ICCS*, Springer LNAI 2746, pp. 160-173, 2003.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 71-76.
ISBN 978-88-8303-295-0. EUT, 2010.

Identification with iterative nearest neighbors using domain knowledge

David Grosser, Noël Conruyt, Henri Ralambondrainy

Abstract — A new iterative and interactive algorithm called CSN (Classification by Successive Neighborhood) to be used in a complex descriptive objects identification approach is presented. Complex objects are those designed by experts within a knowledge base to describe taxa (monography species) and also real organisms (collection specimens). The algorithm consists of neighborhoods computations from an incremental basis of characters using a dissimilarity function which takes into account structures and values of the objects. A discriminant power function is combined with domain knowledge on the features set at each iteration. It is shown that CSN consistently outperforms methods such as identification trees and simplifies interactive classification processes comparatively to search for K-Nearest-Neighbors method.

Index Terms — identification, Similarity, K-Nearest-Neighbors, Decision Trees, structured data, knowledge base, life science.

1 Introduction

In the frame of environmental sciences, for helping to preserve rich ecosystems from biodiversity loss, the acquisition and production of knowledge on biological specimens and taxa is an essential part of the work of systematicians [1]. Indeed, being able to describe, classify and identify a specimen from morphological characters is a first step for monitoring biodiversity, because it gives access to information relative to its species name (Biology, Geography, Ecology, Taxonomy, bibliography, photography). These tasks can be assisted in biodiversity informatics by databases for storing information and computer science decision support tools for description, classification and identification purpose with knowledge bases. In return, these complex representations deliver interesting models and processing problems to deal both with *domain knowledge* and specimen descriptions.

In many fields of real world applications, we can capture a given aspect of the descriptive domain knowledge by associating attributes of the problem

The authors are with the Computer Science and Mathematics laboratory (LIM-IREMIA) of Reunion University – 97400 Saint-Denis, France. E-mail: grosser@univ-reunion.fr.

structure with objects linked by composition and/or specialization relationships. We can also structure the domain definition of nominal attributes by a hierarchy of values. These techniques enable the algorithms to take into account mutual dependencies between attributes and values and to compare case properties with more accuracy.

For instance, for the knowledge base on Corals of the Mascarene Archipelago (http://coraux.univ-reunion.fr/), the descriptions of specimens are often highly structured (composite objects, taxonomic attributes), highly noisy (erroneous or unknown data) and polymorphous (variable, i.e. simultaneous presence of states or imprecise data). To take into account this complexity, we need to define a *descriptive model* (*or Ontology)* that includes information about objects' relationships, attribute types and other semantic aspects: scope of the values, meaning of special values (defaults, exceptions), observation cost of characters.

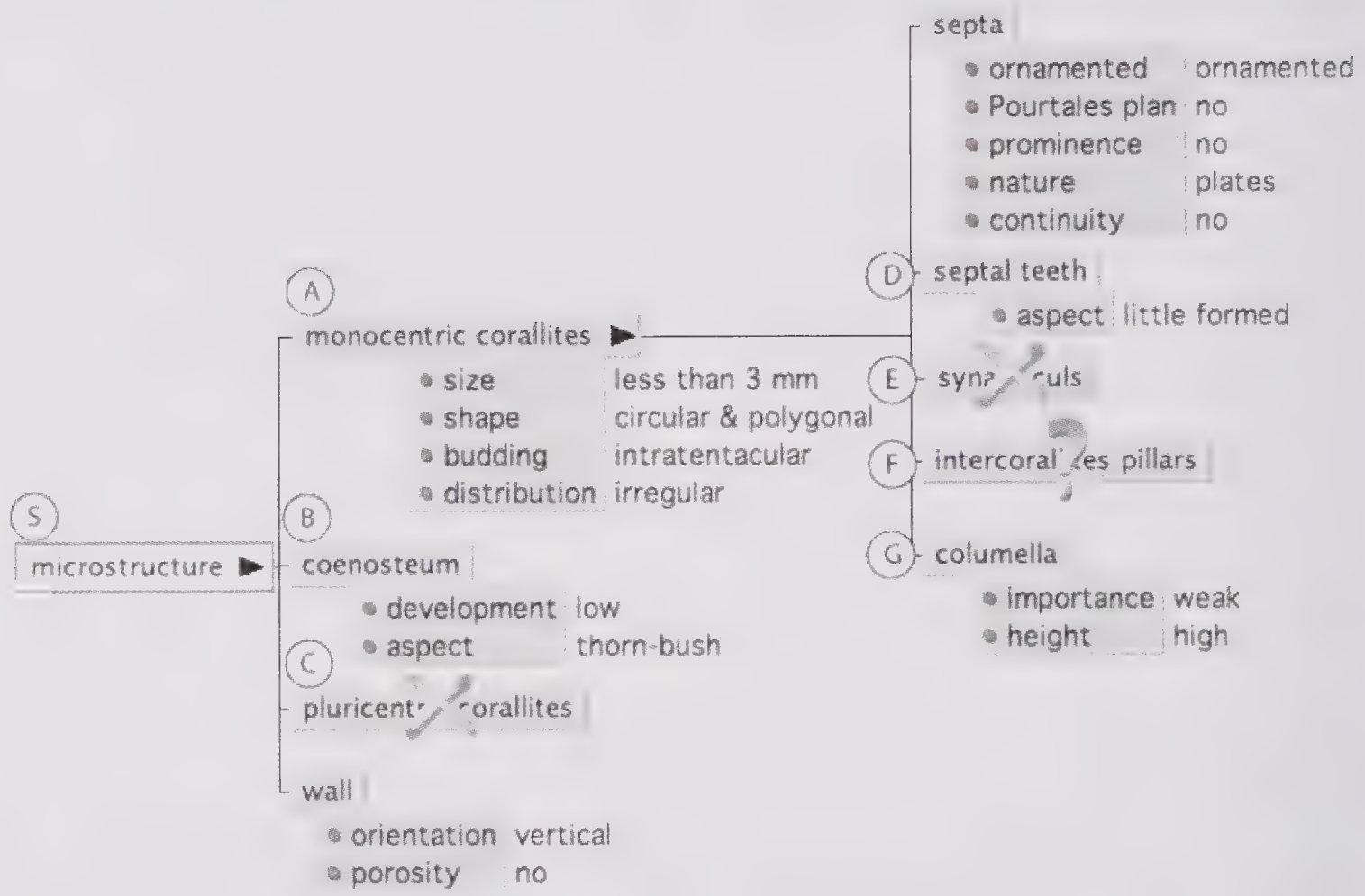


Fig. 1 – Part of a specimen description made with IKBS. Characters (attributes) are attached to objects (eventually missing or absent) that are organized with composition relationships.

For engineering Systematics, we have developed a type of knowledge base system that supports both taxa and specimens descriptions. IKBS (Iterative Knowledge Base System) is a knowledge management system available on the Internet (http://ikbs.sourceforge.net) that helps to define descriptive models, describe instances of these models (see Fig. 1) and then identify new specimen descriptions with different identification methods: an Identification Tree based method (monothetic) and a K-Nearest Neighbors method (polythetic) that uses a dissimilarity function designed to deal with such complex objects representations [2].2 Classification by Successive Neighborhood (CSN)

2 Classification by Successive Neighborhood (CSN)

CSN is a new iterative and interactive method that uses a similarity measure and a discriminant character selection to identify complex objects [3]. Starting from a partially described unlabeled object, the method consists in selecting at each step an objects' neighborhood in regards to a similarity measure. A set of candidate classes is computed from the neighbor set considering class frequencies. A list of discriminant characters is built from the neighborhood and the best is chosen among that list. The value is obtained interactively from users (or another data source). A new neighborhood is computed on the basis of the new partial description of the object. The process iterates until the candidate classes set is homogeneous.

The iterative process to identify an unlabeled description called e is made of the following functions:

2.1 Building neighbors set

The neighbors of e at iteration m is the set V_m of objects inside of the sphere of radius Δ_m centered at e:

$$V_m = \{o \in O \mid d(e,o) < \Delta_m\},\ d \text{ is a dissimilarity function.}$$

The radius value is determined from the maximum distance, the *max* dissimilarity value between e and the set V_{m-1}:

$$\Delta_m = max\,(d(e, o_i)),\ o_i \in V_{m-1}.$$

Then $\{\Delta_m\}$ is a decreasing sequence.

2.2 Selecting discriminant characters

An ordered list of informative variables is computed at each step of the classification process from V. The first element is exposed as a question to the user who can choose an alternative variable from the list. The list is built in function of three criteria:

- Discriminant power. Choice of different classical criteria computing the information gain used in machine learning such as Shannon entropy measure or Gini index. Straightforwardly, this type of criterion minimizes the number of questions.
- Selectable characters. The method considers only at each step characters that may be indicated for choice. Relative questions about presence or absence of components are also considered as selectable attributes.
- Using attributes weighting in the data model reflecting observation cost or other strategic knowledge about characters.

3 Experiments

In the following experiments, we have extracted some descriptions from the *Fungiidae* Knowledge base on Corals of the Mascarene Archipelago that counts approximately 150 classes and 800 complex objects. We follow a double

objective. Firstly, we aim to illustrate the execution of the CSN algorithm in the IKBS software. Secondly, we want to compare the classification (identification) accuracy of the CSN method in regards to an identification tree (IT) based method and a simple K-Nearest-Neighbors (KNN). Both methods already exist in IKBS and use respectively the same discriminant character selection method and the same dissimilarity function.

3.1 Execution of CSN algorithm

The example in Tab. 1 illustrates the identification process of an unlabeled description *e* initially empty. The *Fungiidae* Knowledge base is made of 63 cases with 94 characters and 15 species names. To simulate user interactions, the data source of the description *e* to identify is set to e_r corresponding to a complete and referenced description of a specimen pertaining to species *Fungia concinna* (case number 8 among 63). For the correct identification, e_r class (species name) is compared with *e* class at the end of the process. The criterion used for the character selection function is the classical Shannon's information gain measure.

Num	Attribute	Value	Neighbors set			
			case	class	value	Distance
1	form[colony]	circular	62	fungites	circular	0.2216
	gain = 0.628		46	fungites	circular	0.2239
			51	scruposa	circular	0.2259
7	size[skeleton]	10.3 cm	62	fungites	9.5 cm	0.1905
	gain = 0.423		46	fungites	9 cm	0.1930
			60	fungites	15.5 cm	0.1974
8	density[spines]	[10 16]	62	fungites	[6 10]	0.1912
	gain = 0.423		46	fungites	6	0.1946
			60	fungites	[6 8]	0.1978
9	profile[skeleton]	flat	60	fungites	flat	0.1924
	gain = 0.394		62	fungites	convex	0.1935
			46	fungites	convex plane	0.1968
19	dev[coenosteum]	sub equal	46	fungites	sub equal	0.1901
	gain = 0.211		**39**	**concinna**	sub equal	0.1915
			60	fungites	not equal	0.1930
20	lobes	absent	46	fungites	absent	0.1901
	gain = 0.173		**39**	**concinna**	absent	0.1915
			60	fungites	absent	0.1930
21	form[spines]	cylindrical	39	**concinna**	cylindrical	0.1912
	gain = 0.153		46	fungites	conical	0.1935
			62	fungites	cylindrical	0.1941

Tab. 1 – Example of identification process by successive iterations (Num) of e.

Tab. 1 shows a selected subset of iterations that conducted to a good detection of *e*. 21 iterations (and so 21 character values) were necessary. For each line, the selected character, the corresponding value and the information gain associated are showed. For each neighborhood, information about the 3 first objects in *V* is shown: cases indexes (in the case base), attaching classes,

values for the selected character and the dissimilarity values. For convenience needs, the stopping criterion used is the exact matching with the class of the first case (in bold in the table).

The most interesting information to observe is the progression of *V*. Variations of positions show how supplying information to *e* modifies distances and consequently the order of cases in *V*. Thus, for instance, between iteration 8

Bases	nb clas	nb case	nb attr	Tree ident		K-neighbors		CSN	
				score	rate	score	rate	score	rate
Faviinae	36	92	146	65	70.65%	85	92.39%	84	91.30%
Montastreinae	15	24	118	17	70.83%	22	91.66%	19	79.16%
Fungiidae	15	63	94	47	74.60%	58	92.06%	55	87.30%
Mussidae	15	56	28	49	87.50%	54	96.42%	51	91.07%
Poritidae	28	28	87	22	78.57%	24	85.71%	19	67.85%
Siderastreidae	14	60	99	49	81.67%	56	93.33%	57	95.00%

Tab. 2 – Comparison of IT, KNN and CSN on classification accuracy of 6 knowledge bases.

and 9, the case 60 goes up to first position because the value of the character *profile of the Skeleton* (noted *profile[Skeleton]*) corresponds to the reference value, but not cases 62 and 46. At iteration 19 appears the case 39 in position 2, labeled with the “good class”. To finish, at iteration 21, the case 39 reached the first position in front of the case 46 and the process stop with a “good matching”.

3.2 Classification accuracy

In this second experiment, we evaluate relative performances of CSN comparatively to IT and to KNN methods already implemented in IKBS. The first IT method [4] is an extension of the supervised classification algorithm C4.5 [5] adapted to the use of a structured descriptive model.

The second KNN method uses the same dissimilarity measure as CSN for its neighbors set computation. The validation method used is a “leave-one-out” cross validation process [6] that consists to classify each case of the base using the others as training set. This method is applied for the three algorithms with similar conditions. For convenience needs, K is set to 1 in the KNN method. Tab. 2 gives results of the validation process on six family knowledge bases. For each base, we show the number of cases and characters, and for each method, identification accuracy (score).

It demonstrates that IT method presents a low accuracy rate comparatively to CSN and KNN. Identification errors are frequents, from 12.5% for *Mussidae* family to 29.35% for *Faviinae*. The best method is KNN that may be intuitively explained by the fact that it uses the overall information of the objects: the reference case is fully described. CSN offers an intermediate score, near KNN and often ouperforms IT. We may observe for instance results of the *Faviinae* base that show a difference among 20% of good identification.

4 Conclusion

To identify a biological object and to associate a class to it, experts usually proceed with two phases. The synthetic phase, by global observation of the most visible characters reduces the field of investigation. The analytical phase, by precise observation of discriminating attributes refines research until obtaining the result. Even if the k-nearest-neighbors approach gives a good classification rate, it is difficult to use in real conditions without background knowledge of the domain. In fact, it is very useful to dispose of an interactive process to design features selection such in decision tree approaches.

The classification by successive neighborhood (CSN) method that we proposed deals with structured and partial objects descriptions. It presents the interest to correspond to the reasoning followed by biologists. Starting from a partial description generally containing the most visible or easy to observe and describe features, the method suggests relevant informa- tion necessary to supplement to determine the most probable class.

We expect that the CSN method is generic and applicable on any fields where structured or semi structured data are considered, such as XML data format or RDF and OWL graph structures. It's enough to lay out a similarity index and a discriminant power function adapted to the considered data.

References

[1] J. E. Winston, *Describing Species: Practical Taxonomic Procedure for Biologists.* New York. Columbia University Press, 1999.

[2] D. Grosser, J. Diatta and N. Conruyt, "Improving dissimilarity functions with domain knowledge". *Proc of the 4th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD'2000)*, pp. 409-415, 2000.

[3] D. Grosser, H. Ralambondrainy and N. Conruyt, "Classification by successive neighborhood". In *KDIR 2009, International Conference on Knowledge Discovery and Information Retrieval*. INSTICC Press, 2009.

[4] N. Conruyt and D. Grosser, "Knowledge engineering in environmental sciences with ikbs". *AI Communications, The European Journal on Artificial Intelligence*, 16(3), pp. 267-278, 2003.

[5] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Series in Machine Learning, 1993.

[6] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection". *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (Morgan Kaufmann, San Mateo), 2 (12), pp. 1137–1143, 1995.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 77-82.
ISBN 978-88-8303-295-0. EUT, 2010.

A MediaWiki implementation of single-access keys

Gregor Hagedorn, Bob Press, Sonia Hetzner,
Andreas Plank, Gisela Weber, Sabine von Mering,
Stefano Martellos, Pier Luigi Nimis

Abstract — Design principles, features, and data exchange options of an open source implementation of single access keys for the collaborative MediaWiki platform are presented.

Index Terms — field guides, flora, fauna, identification tools, *KeyToNature*, social software, SDD, MediaWiki, single-access key, dichotomous key, polytomous key.

1 Introduction

Among the various forms of computer-aided identification keys (compare [1], [2]), single-access keys have long been neglected, being studied primarily as a printable output from character × taxon matrices. However, single-access keys may also be used interactively. Examples are Lucid Phoenix [3] (commercial), Frida/Dryades [4] (closed source), and the two *KeyToNature* open source projects: Open Key Editor [5] and biowikifarm [6] "Wiki keys". The latter implementation, based on JavaScript enhanced MediaWiki [7] templates, is presented here in detail. Its strength is the integration into the collaborative MediaWiki software with much broader applicability for developing floras, faunas or field guides.

2 Design principles

Much of the strength of single access keys derives from information reduction. The information which is actually used in the key (and which must be understood by the user) is only a fraction of the total information present in the descriptions of the taxa.

G. Hagedorn, A. Plank, G. Weber, S. v. Mering are with the Julius Kühn-Institute, Federal Research Centre for Cultivated Plants, Inst. for Epidemiology and Pathogen Diagnostics, Königin-Luise-Str. 19, D-14195 Berlin, E-mail: gregor.hagedorn@jki.bund.de. – Bob Press is with the Natural History Museum London – S. Hetzner is with the Inst. f. Lern-Innovation, Universität Erlangen-Nürnberg, D-91052 Erlangen – S. Martellos, P. L. Nimis are with the Dept. of Life Sciences, University Trieste, I-34127, Trieste, Italy. E-mail: nimis@units.it, martelst@units.it.

In a perfect world with an unlimited number of characters that are convenient and reliable to observe, monomorphic (not variable within species), available for observation at all times, and splitting the remaining taxa into evenly sized partitions, the most reduced identification keys would be the best. In the real world, however, the various imperatives for character selection are in conflict. The resulting keys often use character combinations instead of single characters and may provide verification characters that are not strictly necessary. Similarly, several suboptimal illustrations may be necessary to understand character variability.

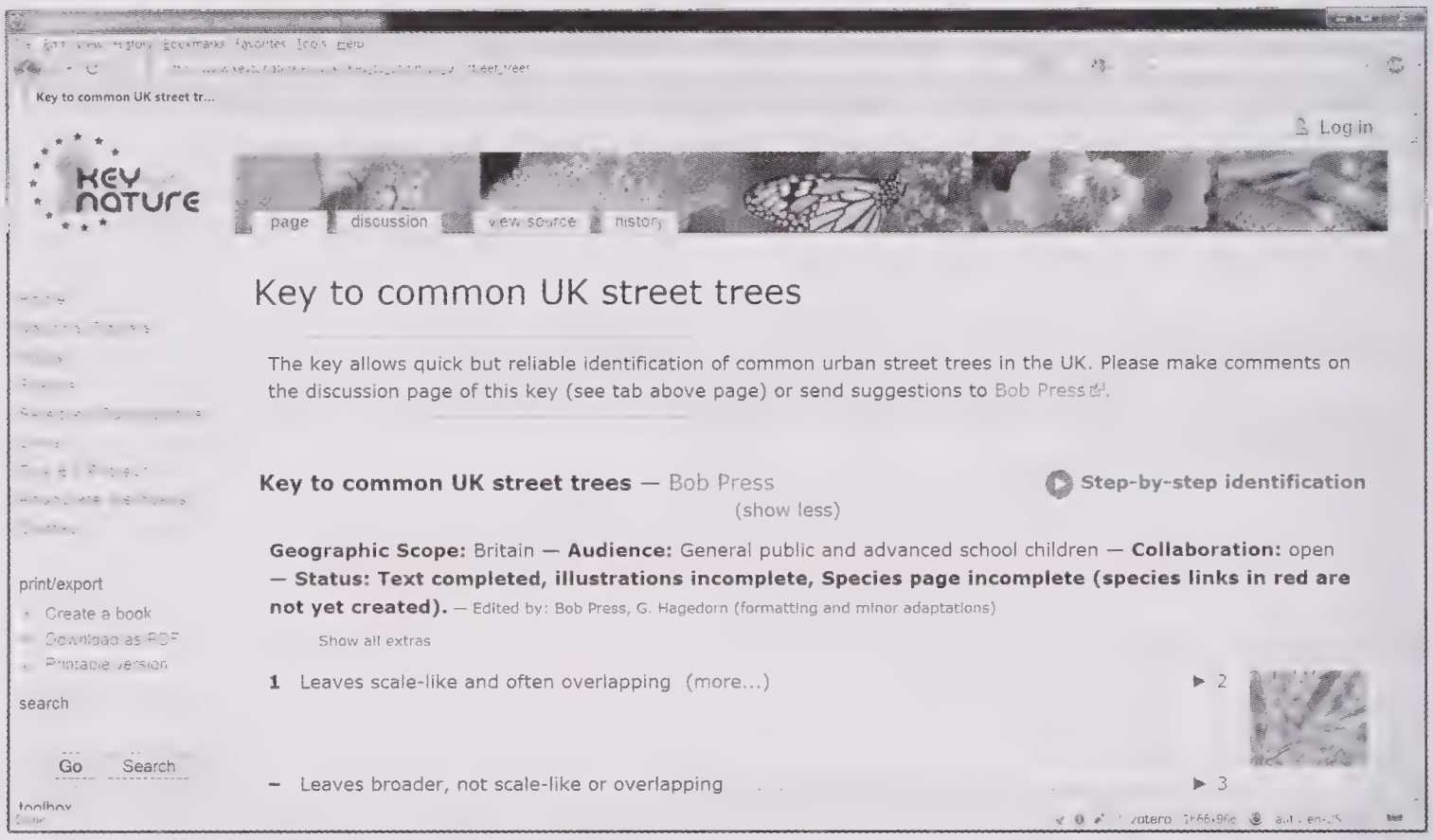

Key to common UK street tr...
Log in
page
discussion
view source
history
Key to common UK street trees
The key allows quick but reliable identification of common urban street trees in the UK. Please make comments on the discussion page of this key (see tab above page) or send suggestions to Bob Press.
Key to common UK street trees — Bob Press
(show less)
Step-by-step identification
Geographic Scope: Britain — Audience: General public and advanced school children — Collaboration: open — Status: Text completed, illustrations incomplete, Species page incomplete (species links in red are not yet created). — Edited by: Bob Press, G. Hagedorn (formatting and minor adaptations)
Show all extras
print/export
Create a book
Download as PDF
Printable version
search
Go
Search
toolbox
1 Leaves scale-like and often overlapping (more...)
▶ 2
– Leaves broader, not scale-like or overlapping
▶ 3

Fig. 1 – A Wiki key [8] as part of a wiki page, presented in overview mode ("Step-by-step identification" starts the interactive mode), and with part of the information hidden ("more…" will display the hidden information). The key metadata (Geographic Scope following) are initially displayed, but hideable.

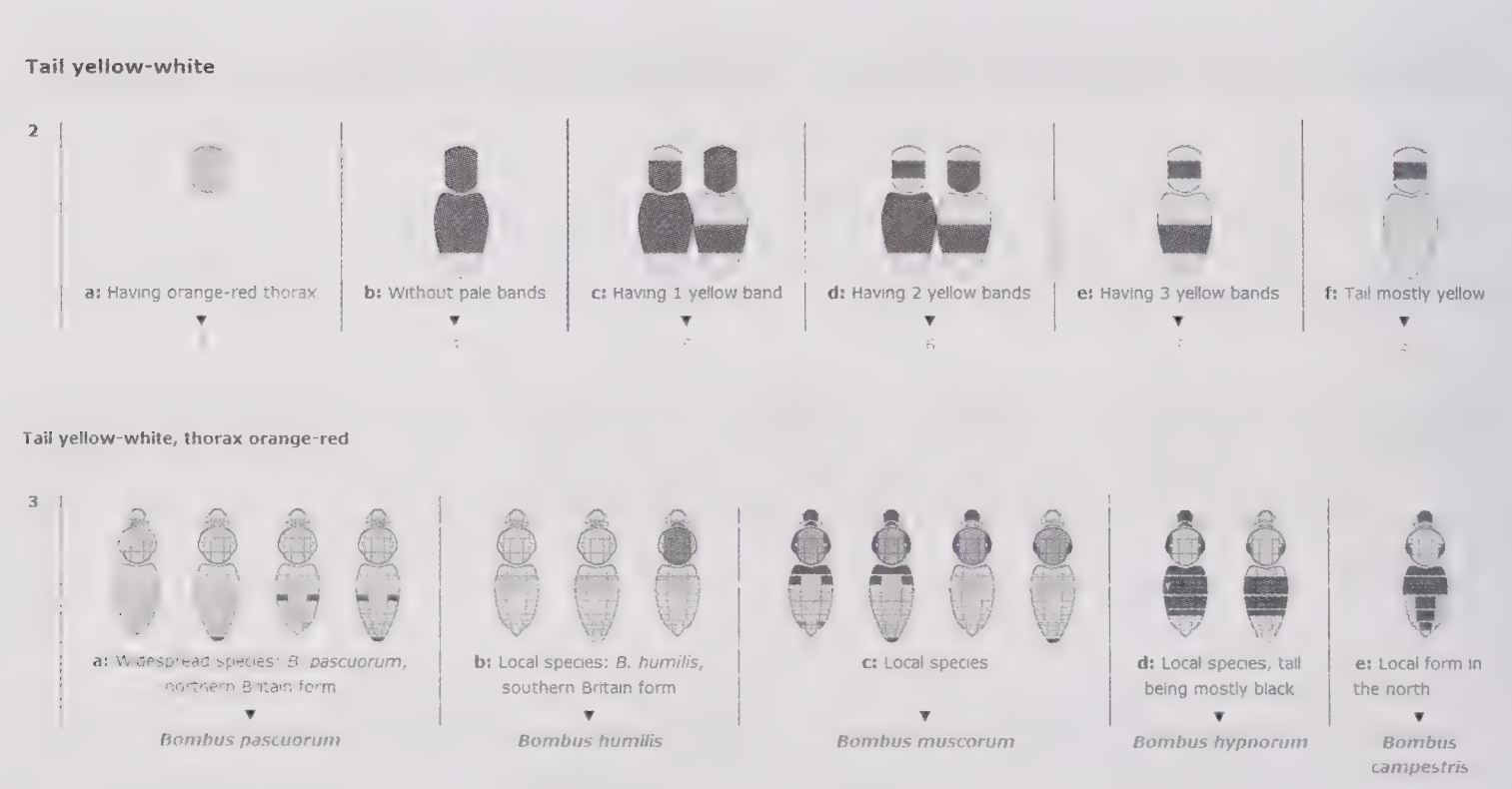


Fig. 2 – A variant wiki template generates a horizontal (side-by-side) layout. Both the horizontal and vertical layout can be used interactively and both provide for sub-headings ("Tail yellow-white", etc.).

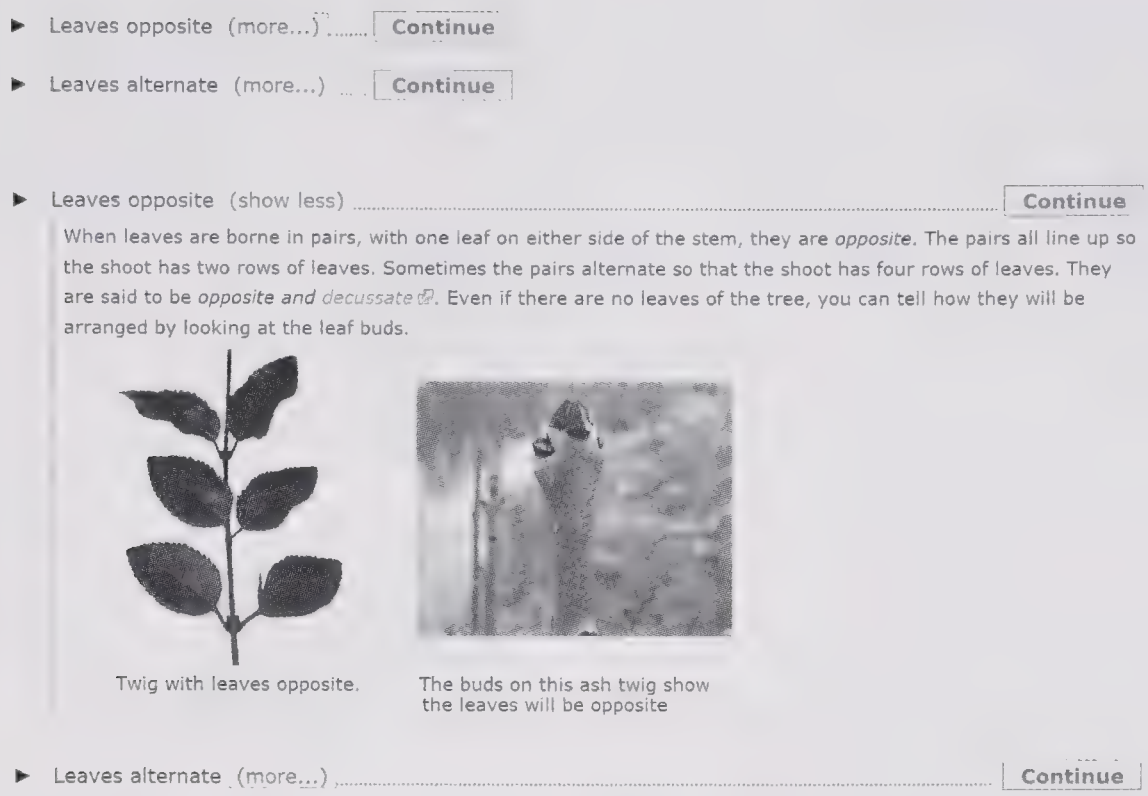

► Leaves opposite (more...) Continue
► Leaves alternate (more...) Continue
► Leaves opposite (show less) Continue
When leaves are borne in pairs, with one leaf on either side of the stem, they are opposite. The pairs all line up so the shoot has two rows of leaves. Sometimes the pairs alternate so that the shoot has four rows of leaves. They are said to be opposite and decussate. Even if there are no leaves of the tree, you can tell how they will be arranged by looking at the leaf buds.
Twig with leaves opposite.
The buds on this ash twig show the leaves will be opposite
► Leaves alternate (more...) Continue

Fig. 3 – Extra information is initially hidden (top). It will be displayed (bottom) after clicking on "more…".

In the MediaWiki single-access keys, initially only terse lead statements (which should be limited to information strictly complementary within a couplet) and a subset of images are shown (in standard layout, Fig. 1, small previews in a margin plus up to 2 larger images below each lead; in horizontal layout, Fig. 2, up to 8 images above the lead statement).

As a form of "information reduction management", further information is shown only on request: general remarks, a gallery of up to 8 secondary illustrations, and – for taxa – synonyms, brief diagnosis, occurrence, distribution, and habitat information (Fig. 3).

All images are displayed at a reduced size, sufficient only to enable the user to decide if enlarging is worthwhile. The zoom functionality acts in the page context (Fig. 4). It further provides a link to creator, copyright, and license information, required to be directly accessible by many licenses.

Information hiding is implemented through JavaScript in a graceful degradation design (i. e. without JavaScript, no information is hidden).

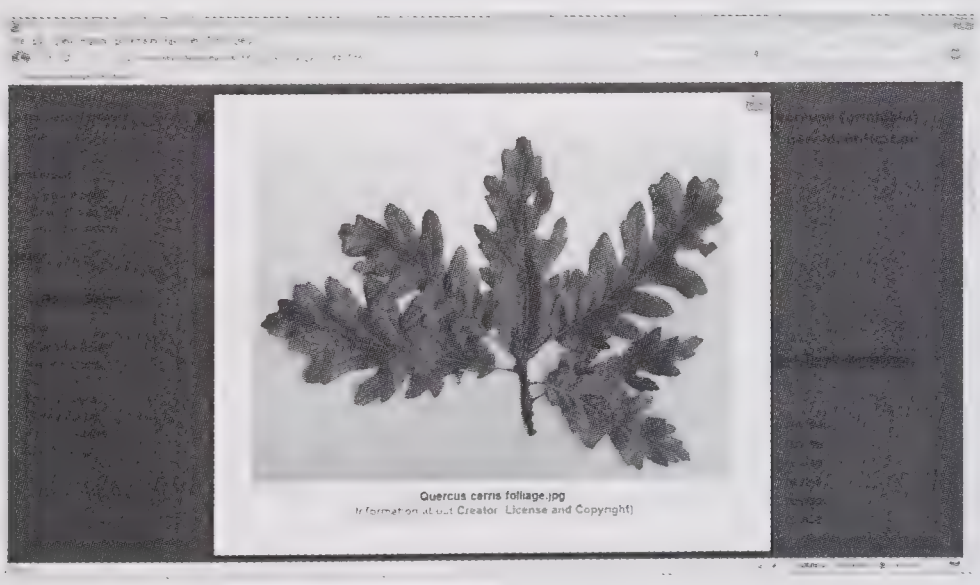


Fig. 4 – Enlarged image as an overlay to the page. At the bottom an image caption and a link to the IPR information (legally required by the Creative Commons licenses) is provided.

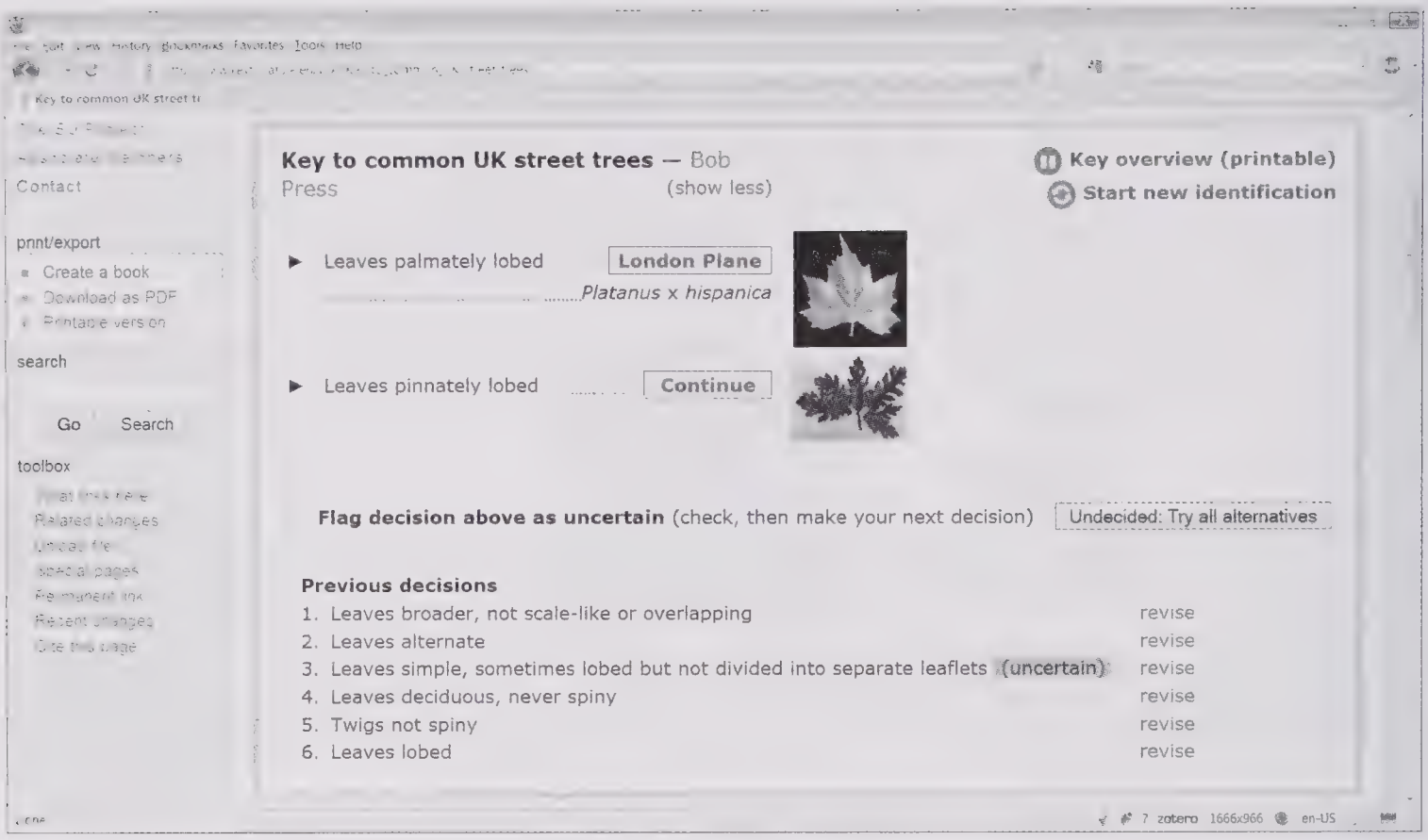


Fig. 5 – Wiki key in JavaScript-based interactive mode. The history of previous decisions is displayed at the bottom, with decision 3 having been marked as uncertain.

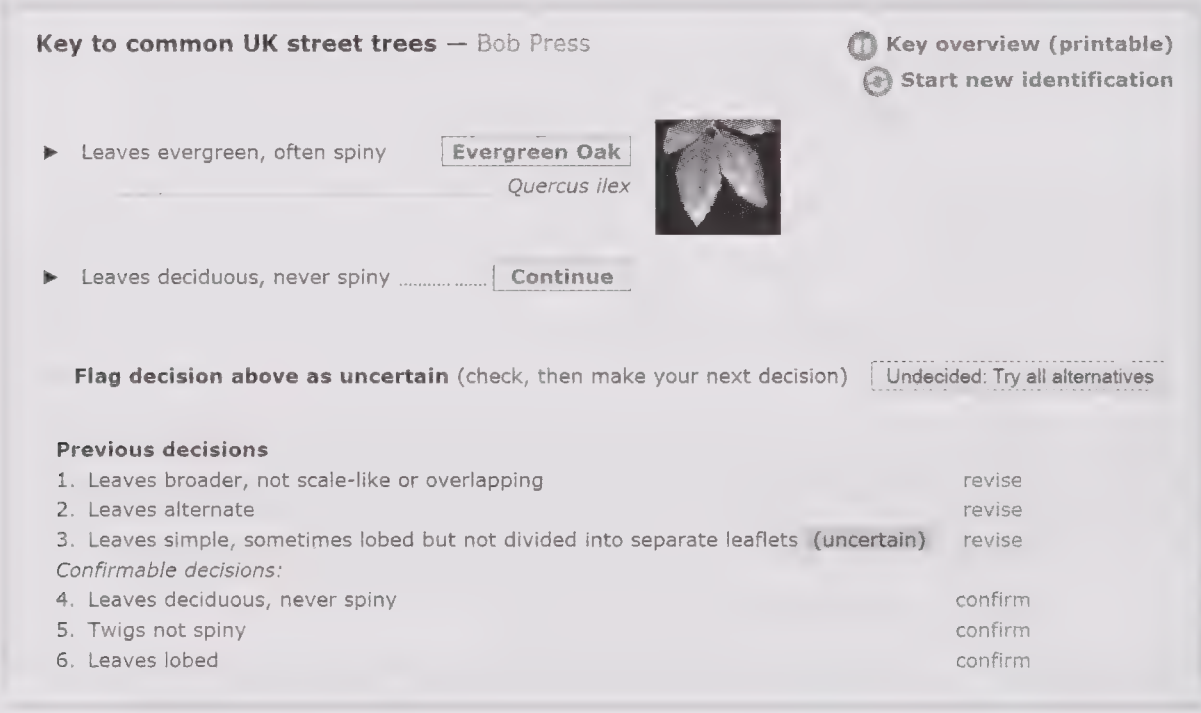

Key to common UK street trees — Bob Press
Key overview (printable)
Start new identification
Leaves evergreen, often spiny
Evergreen Oak
Quercus ilex
Leaves deciduous, never spiny
Continue
Flag decision above as uncertain (check, then make your next decision)
Undecided: Try all alternatives
Previous decisions
1. Leaves broader, not scale-like or overlapping revise
2. Leaves alternate revise
3. Leaves simple, sometimes lobed but not divided into separate leaflets (uncertain) revise
Confirmable decisions:
4. Leaves deciduous, never spiny confirm
5. Twigs not spiny confirm
6. Leaves lobed confirm

Fig. 6 – Steps in the history or previous decisions are revisable (here step 4 is being revised). Later steps are confirmable and discarded only if a conflicting decision is taken.

3 Interactive step-by-step identification

The printable overview can be switched to an interactive mode. The open source jKey player [9] then displays only a single choice at a time (Fig. 5), helping esp. students to concentrate. To maintain context, a history of previous decisions is provided. Steps are revisable, maintaining progress as long as possible (Fig. 6). The history also helps to verify identifications, since it can be read as a description of the object that is being identified.

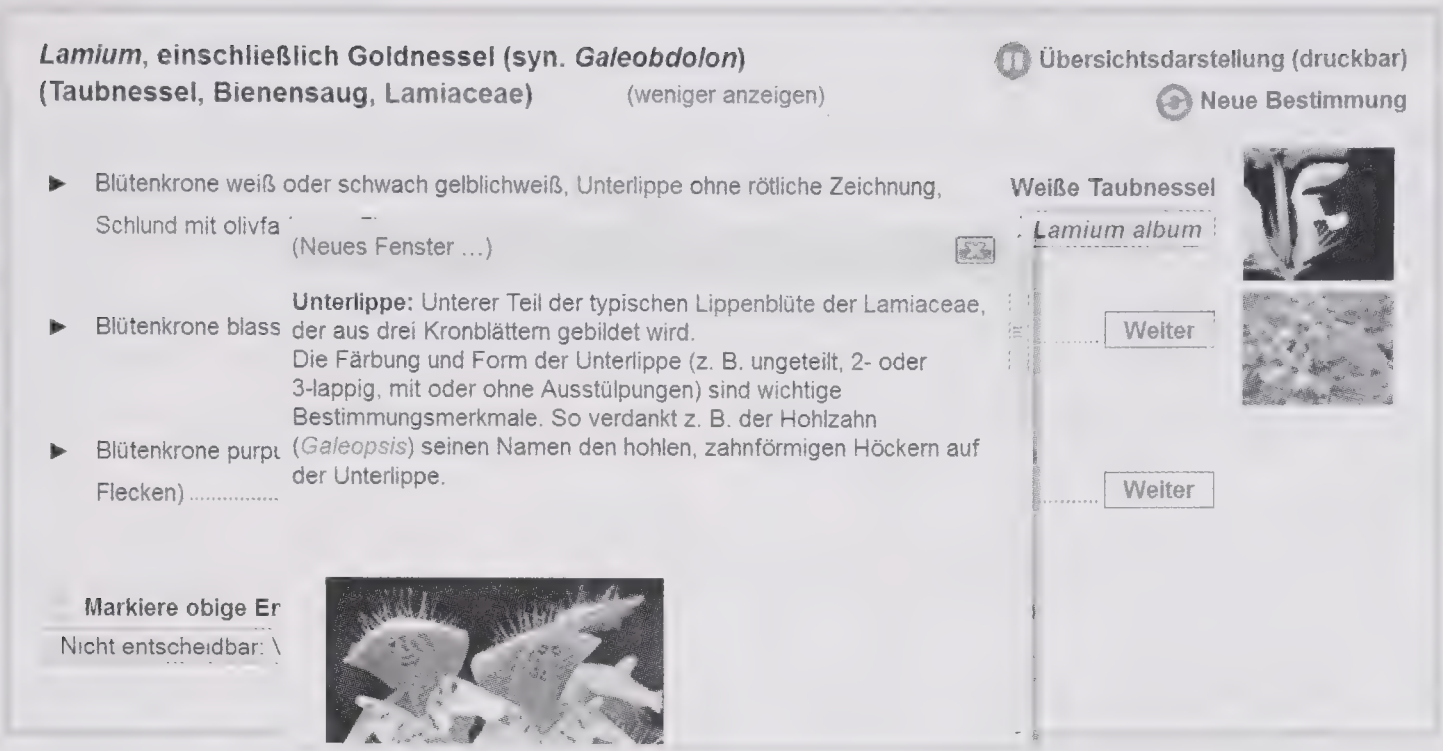

Lamium, einschließlich Goldnessel (syn. Galeobdolon)
(Taubnessel, Bienensaug, Lamiaceae)
(weniger anzeigen)
Übersichtsdarstellung (druckbar)
Neue Bestimmung
Blütenkrone weiß oder schwach gelblichweiß, Unterlippe ohne rötliche Zeichnung, Schlund mit olivfa
(Neues Fenster ...)
Blütenkrone blass
Blütenkrone purpu Flecken)
Unterlippe: Unterer Teil der typischen Lippenblüte der Lamiaceae, der aus drei Kronblättern gebildet wird.
Die Färbung und Form der Unterlippe (z. B. ungeteilt, 2- oder 3-lappig, mit oder ohne Ausstülpungen) sind wichtige Bestimmungsmerkmale. So verdankt z. B. der Hohlzahn (Galeopsis) seinen Namen den hohlen, zahnförmigen Höckern auf der Unterlippe.
Weiße Taubnessel
Lamium album
Weiter
Weiter
Markiere obige Er
Nicht entscheidbar:

Fig. 7 – Display of a glossary definition in an overlay on the web page (not in a separate window).

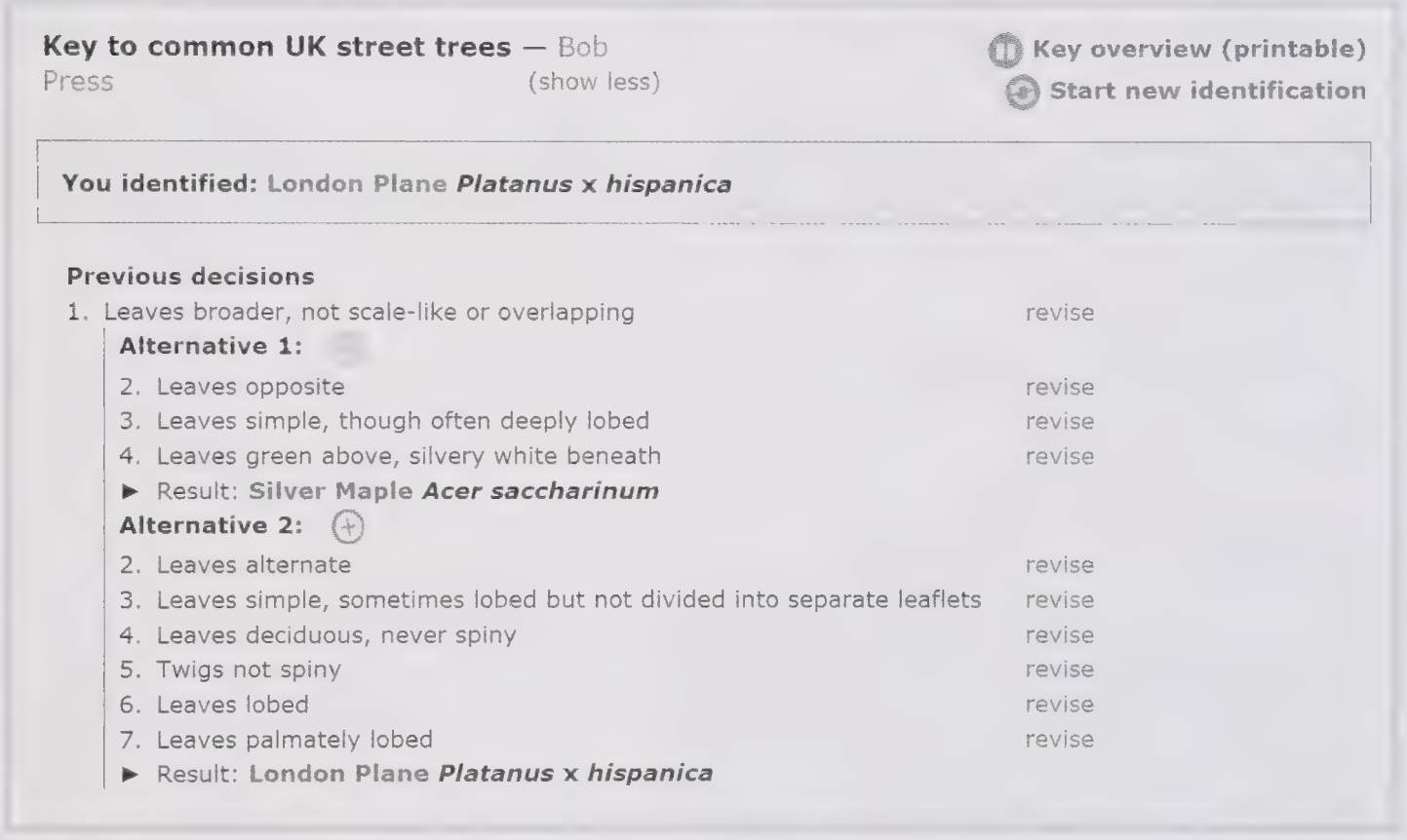

Key to common UK street trees — Bob
Press
(show less)
Key overview (printable)
Start new identification
You identified: London Plane Platanus x hispanica
Previous decisions
1. Leaves broader, not scale-like or overlapping revise
Alternative 1:
2. Leaves opposite revise
3. Leaves simple, though often deeply lobed revise
4. Leaves green above, silvery white beneath revise
► Result: Silver Maple Acer saccharinum
Alternative 2:
2. Leaves alternate revise
3. Leaves simple, sometimes lobed but not divided into separate leaflets revise
4. Leaves deciduous, never spiny revise
5. Twigs not spiny revise
6. Leaves lobed revise
7. Leaves palmately lobed revise
► Result: London Plane Platanus x hispanica

Fig. 8 – After selecting the button: “Undecided: try all alternatives”, multiple alternative paths may be followed.

4 Data exchange

The *KeyToNature-Dryades*/Frida system provides a special export format that directly creates text formatted to be pasted into wiki pages as ready wiki text. Wiki keys can further be converted to SDD xml data by a converter created at the Natural History Museum in London.

5 Pedagogical features

The single-access keys are supported by several pedagogically relevant features:

1. Illustrated concept definitions and help pages, stored as editable wiki pages, may be accessed from any point in the identification key, providing context-sensitive help. When the user hovers the mouse over a term, the definition opens in a pop-up layer (Fig. 7). From there a new window can be opened if needed.

2. The history allows direct access to and revision of any previous decision. This may occur in a dialogue with the teacher, who can help in reviewing misinterpretations.

3. Users can flag particular decisions as "uncertain" (Fig. 5). Although this does little more than marking a step in the history, it can greatly enhance the student-teacher communication. It allows students to actively seek teacher assistance at a time when he or she is available, while continuing their work in the meantime.

4. The interactive mode offers an option to not take a decision at a given step, allowing users to explore the key in multiple directions. After selecting the "Undecided: try all alternatives" button, the player will continue with the first alternative. However, the history allows the user to switch between alternative branches, recording in all branches (Fig. 8).

Acknowledgement

This work was supported by the *KeyToNature* Project, ECP-2006-EDU-410019, in the *e*Content*plus* Programme.

References

[1] R. J. Pankhurst, *Practical Taxonomic Computing*, 1991.

[2] G. Hagedorn, G. Rambold and S. Martellos 2010, "Types of identification keys". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 59-64, 2010.

[3] Lucid Phoenix (http://www.lucidcentral.org/LinkClick.aspx?link=152), 2010-07.

[4] S. Martellos and P. L. Nimis, "*KeyToNature*: Teaching and Learning Biodiversity. Dryades, the Italian Experience." In: M. Muñoz, I. Jelìnek, F. Ferreira (eds.), *Proceedings of the IASK International Conference Teaching and Learning 2008*, pp. 863-868, 2008.

[5] S. Martellos, E. v. Spronsen, D. Seijts, N. Torrescasana Aloy, P. Schalk and P. L. Nimis, "User-generated content in the digital identification of organisms: the *KeyToNature* approach" *Int. J. Information and Operations Management Education*, vol. 3, 3, pp. 272-83, 2010.

[6] MediaWiki software, http://www.mediawiki.org/wiki/MediaWiki, 2010-07.

[7] G. Hagedorn, G. Weber, A. Plank, M. Giurgiu, A. Homodi, C. Veja, G. Schmidt, P. Mihnev, M. Roujinov, D. Triebel, R. A. Morris, B. Zelazny, E. van Spronsen, P. Schalk, C. Kittl, R. Brandner, S. Martellos and P. L. Nimis, "An online authoring and publishing platform for field guides and identification tools". In: Nimis P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 13-18, 2010.

[8] B. Press, Key to common UK street trees. http://www.keytonature.eu/wiki/Key_to_common_UK_street_trees, 2010-07.

[9] S. Opitz and G. Hagedorn, "The jKey wiki key player and builder". *Proc. of TDWG 2009*, pp. 9-13 Nov. 2009, Montpellier, 2009.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 83-88.
ISBN 978-88-8303-295-0. EUT, 2010.

Simple matrix keys from Excel spreadsheets

Gregor Hagedorn, Mircea Giurgiu, Andrei Homodi

Abstract — An innovative workflow is presented, leading from a simple character and taxon matrix, prepared in an Excel spreadsheet, to a functioning free-access matrix key, stored in SDD (Structured Descriptive Data) format and presented in the flash-based free-access key player IBIS-ID directly on the web. The subset of matrix key functionality supported by this workflow include categorical and quantitative characters, multiple character states for a given taxon, taxon, character, and state illustrations, and a single character grouping level. Advanced features such as complex character or taxon trees, character dependency, or taxon-specific state images are not supported. The workflow aims to attract new contributors for which the learning curve of special purpose matrix-key editing software is too steep.

Index Terms — identification tools, free-access key player IBIS-ID, DELTA, SDD.

1 Introduction

Identification tools such as free-access (= multi-access) or multi-entry keys that are based on a character × taxon data matrix (i. e. a table with the taxa in one dimension and characters in the other) have certain advantages over single-access keys [1, 2, 3, 4]. However, creating a computer-aided matrix key typically requires learning a special purpose application. This limits the number of matrix-based keys produced by biologists, who tend to produce keys similar to the single-access keys typically encountered in the printed literature.

Many biologists are acquainted with spreadsheet applications, especially Microsoft Excel, to edit tabular data. Unfortunately, the visualization of a simple character × taxon table – for which spreadsheet applications are ideal – is a simplified idealization of a more complex data model. All of taxa, characters, states, and the descriptive matrix cells may have further structure:

1. Taxa may require common and scientific names, web links to taxon pages, images, brief diagnostic text, etc.
2. Characters may require a data type, a list of supported states (i. e.

G. Hagedorn is with the Julius Kühn-Institute, Federal Research Centre for Cultivated Plants, Institute for Epidemiology and Pathogen Diagnostics, Königin-Luise-Str. 19, D-14195 Berlin, Germany, E-mail: gregor.hagedorn@jki.bund.de – M. Giurgiu and A. Homodi are with the Telecommunications Department, Technical University of Cluj-Napoca, Cluj 400027, Romania. E-mail: Mircea.Giurgiu@com.utcluj.ro.

constraining the vocabulary), illustrations, explanatory notes.

3. States may provide illustrations or explanatory notes.

The cells of the matrix may contain multiple values, modifiers, notes, and taxon-specific state or character illustrations. For a character: “flower colour” the cell content may be: “usually pink, sometimes red, or blue (immediately after opening)”; for a character: “stem hairiness” it may be “long (2-5 mm) or medium long (1-2 mm) hairs”. A character having more than one state in a taxa is called “polymorphic” in biology. It may occur as a result of a true genetic polymorphism in a population, environmentally induced phenotypic variation (e. g., occurring within the set of flowers on a single plant), or relatively minor quantitative variation that happens – in the present taxon – to cross the artificially drawn borders of a continuously varying character (such as hairiness).

However, when relatively simple rules are followed it is possible to support a subset of the potential complexity of matrix keys within spreadsheets nevertheless.

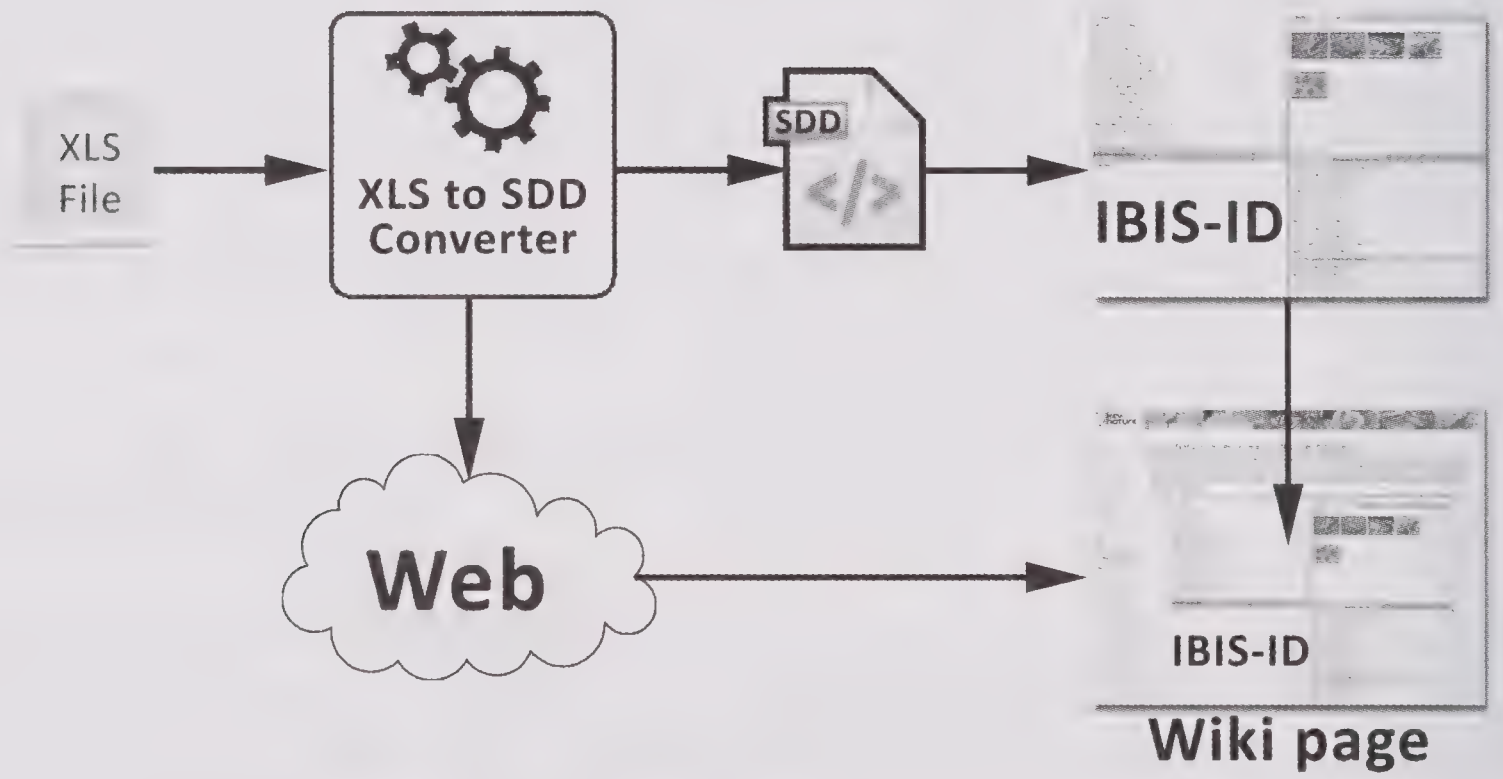


Fig. 1 – The workflow from a Microsoft Excel spreadsheet to SDD conversion and presentation with the free-access key player IBIS-ID inside and automatically published inside a MediaWiki web page.

2 The spreadsheet

The workflow (Fig. 1) starts with the creation of a character × taxon matrix by the biologist, following either instructions on the web [5] or supported by a downloadable template. The simplest layout is indeed one with characters being named in the first row, taxa in the first column and the remainder filled with the taxon × character data (Fig. 2).

For categorical characters (ordinal or nominal scale) the categorical value or state is expressed directly using its label or “name” (e. g., “red”) rather than using a code. In contrast, the DELTA [6] or SDD formats use numeric character and state codes to enforce higher consistency. Multiple states are supported by separating the state names with a semicolon, slash or ampersand (example: “red; blue“). The semicolon is provided as an intuitive delimiter for most

	1	2	3	4	5	6
1	**Name [de]**	**Wissenschaftlicher Na**	**Wuchshöhe [cm]**	**Blattrand**	**Wuchsform**	**Stängel**
2	Kopf-Zwergginster	Chamaecytisus supinu	30-80	ganzrandig [[File:Blat	Kleinstrauch	verholzt, wenig behaa
3	Rauhaar-Zwergginster	Chamaecytisus hirsutu	15-50	ganzrandig	Kleinstrauch	verholzt, wenig behaa
4	Schwarzwerdender Ge	Cytisus nigricans	30-100	ganzrandig	Kleinstrauch	gerillt, behaart
5	Gewöhnlicher Besengi	Cytisus scoparius	-200	ganzrandig	Strauch	kantig, ± kahl
6	Kugel-Ginster	Genista radiata	30-80	ganzrandig	Strauch	anliegend behaart
7	Sichel-Luzerne	Medicago falcata	20-60	vorne gezähnt	aufsteigend-niederlieg	kantig, ± kahl-kraus
8	Hopfenklee	Medicago lupulina	10-30	vorne gezähnt	aufsteigend-niederlieg	kantig, anliegend beha
9	Zwerg-Schneckenklee	Medicago minima	5-30	vorne gezähnt	aufsteigend-niederlieg	dicht anliegend behaa
10	Montpellier-Bockshor	Medicago monspeliac	5-20	spitz gezähnt	niederliegend, einjähri	kurz behaart
11	Saat-Luzerne	Medicago sativa	30-90	vorne gezähnt	aufsteigend-niederlieg	kantig, ± kahl-behaart
12	Weißer Steinklee	Melilotus albus	30-150	± deutlich gezähnt	aufrecht, einjährig	± kurz behaart
13	Hoher Steinklee	Melilotus altissimus	30-120	± scharf gezähnt	aufrecht, einjährig	meist kahl
14	Kleinblütiger Steinklee	Melilotus indicus	15-50	± stumpf gezähnt	aufsteigend-niederlieg	± kahl
15	Echter Steinklee	Melilotus officinalis	30-120	gezähnt [[File:Blattra	aufrecht, einjährig	kahl
16	Gefurchter Steinklee	Melilotus sulcatus	10-50	gezähnt	aufrecht, einjährig	± kahl
17	Gelbe Hauhechel	Ononis natrix	20-40	gezähnt	aufrecht, mehrjährig	rund, drüsig
18	Zwerg-Hauhechel	Ononis pusilla	10-30	gezähnt	aufrecht, mehrjährig	aufrecht
19	Kriechende Hauhechel	Ononis repens	30-60	gezähnt	aufsteigend-niederlieg	grün, allseitig behaart
20	Rundblatt-Hauhechel	Ononis rotundifolia	15-40	gezähnt	aufrecht, mehrjährig	drüsig behaart
21	Dornige Hauhechel	Ononis spinosa s.l.	30-60	gezähnt	aufrecht, mehrjährig	rötlich, 1-2-zeilig beha
22						
23	Title	Schlüssel zu dreiblättrigen Fabaceen (Testdatensatz)				
24	Creators	Die Autorin				
25	Language	de				
26	Sources					
27	Copyright	© A. Autorin				
28	License	Creative Commons by-sa 3.0				

Fig. 2 – A Microsoft Excel spreadsheet with a simple data matrix. Visible are an extra column for scientific taxon names (“Wissenschaftlicher Name” in German), the addition of a measurement unit (“[cm]”) in brackets after quantitative characters, state images in the column “Blattrand”, and the metadata for the entire dataset or identification tool at the bottom.

biologists; the “/” and “&” to help those who also use DELTA tools.

The drawback of the direct use of state labels is that the vocabulary of available states is not controlled. This is a purposeful design decision. While it is possible to devise spreadsheet layouts that include separate state listings, we have noticed that our test users found all options to be too difficult and confusing and were unable to create them autonomously. The vocabulary control is therefore postponed to the publication of a first draft of the identification tool in the IBIS-ID player. In the implemented workflow, the IBIS-ID key player will make undesirable entries (combinations of states with modifiers or spelling variants of states) transparent and users can modify their data for the next revision. “Normalizing” the state labels is well supported by the typical search-and-replace functionality of spreadsheet software. While careful planning and control is essential for large matrix projects covering hundreds of characters and taxa, the workflow presented here aims at smaller datasets, where a post-data-entry-validation workflow may result in more agile contributions than a plan-ahead workflow.

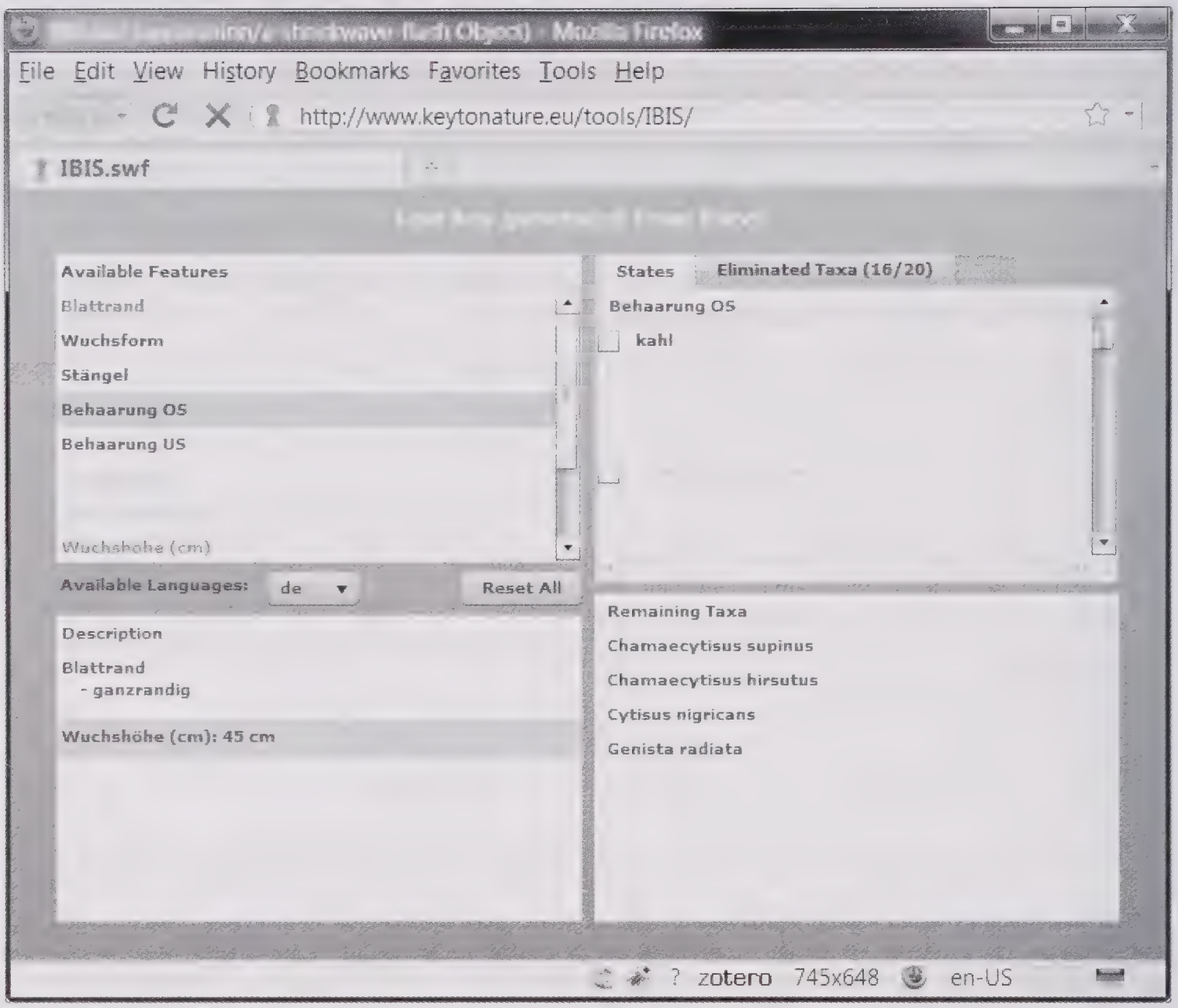


Fig. 3 – The resulting interactive matrix key running under IBIS-ID (here in stand-alone mode, not embedded in a web page).

For quantitative characters a DELTA-like encoding is supported. Various combinations of minimum-maximum (in parentheses), "typical range" and a mean are possible (example: "(1-) 3-6.4-8.2 (-15)").

The characters themselves may be grouped into character groupings by adding the group name surrounded by "{{…}}" after the character label (i.e. in the first row). Presently only a single grouping level is supported.

Matrix cells further support modifiers ("usually", "rarely", "about", "weakly", etc.) and free-form text comments, if they are enclosed in DELTA-like "<…>" markup.

The spreadsheet rules further provide for the inclusion of multiple taxon, character, and state-specific images. Character and state images must be included in double square brackets after the respective label; in contrast taxon images may be placed in an extra column. Not supported are taxon-×-character and taxon-×-state-specific images.

For taxa, several additional column for images, common and scientific names, and web links can be provided. Finally, metadata for the entire key like creators, title, copyright, license, source may be added.

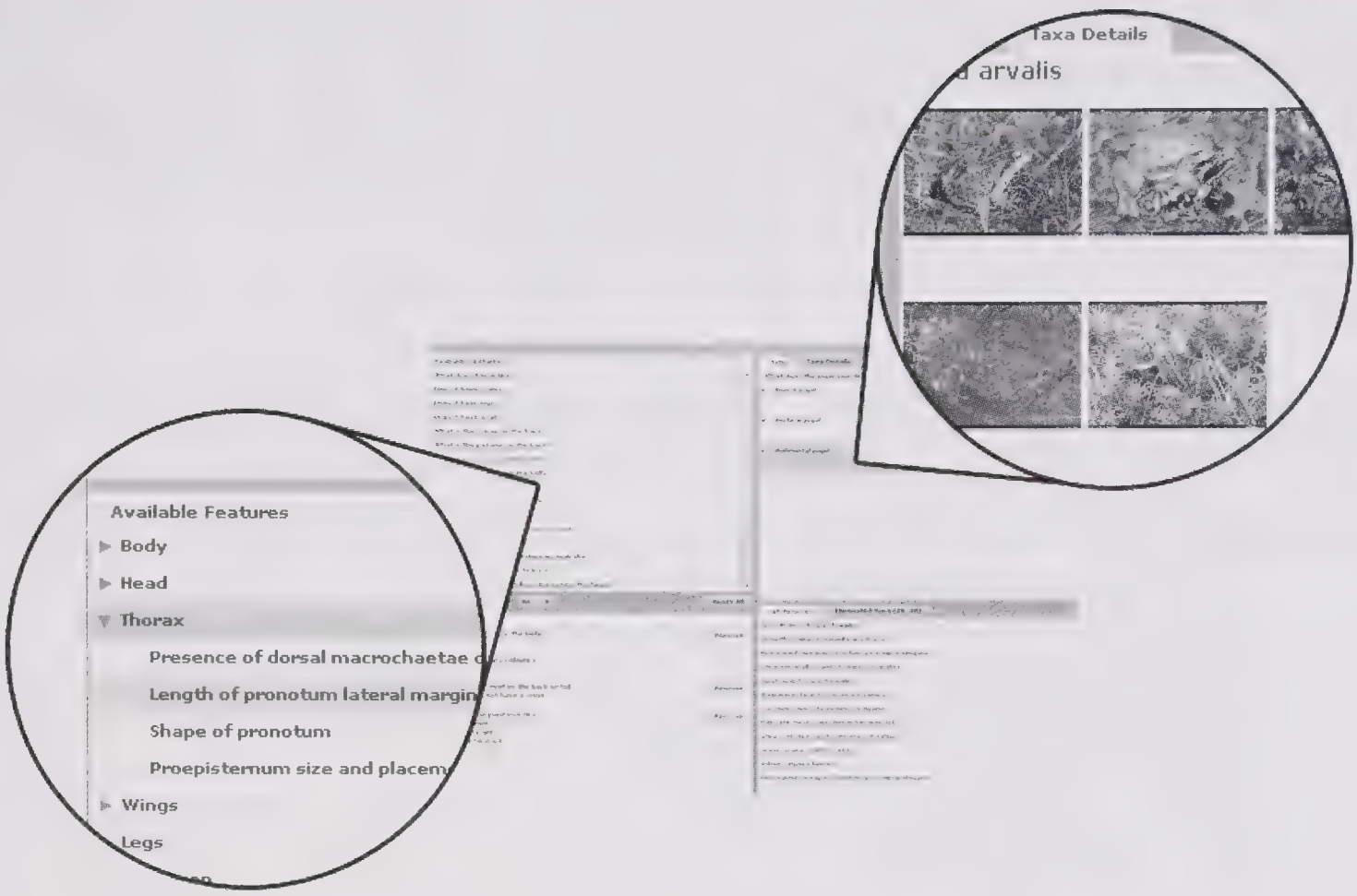


Fig. 4 – Details of IBIS-ID key player, showing character grouping (left) and state images (right).

3 The converter

The converter is presently a downloadable Microsoft .NET for standalone applications (a web-based version is planned). The converter takes the spreadsheet in Microsoft Excel (XLS) format and converts it into SDD.

The converter supports both wiki-style and direct web image references. Uploading images to the wiki allows users to manage their images for both matrix keys, single-access keys and species pages. The simple wiki-style links are automatically translated by the converter into general web links as they are supported by IBIS-ID key player.

If the converter finds unexpected content, it will report this either as warnings (e. g., *"name contains opening double brackets ('[[') but no closing ones, a malformed image may be present"*) or as errors. Error handling is considered important and efforts have been made to help biological users to understand minor errors. If no errors are encountered, the resulting SDD file will be uploaded to a web repository on the MediaWiki based biowikifarm [7].

Furthermore, to enrich the user experience, a wiki page containing the necessary statement to embed the IBIS-ID player [8] (Fig. 3 and 4) inside a wiki page is also generated.

4 Conclusions

It is possible to replace some features that Excel is missing to directly support matrix keys with rules that rely on simple text delimiters. The method is similar to that used by DELTA the special purpose Windows DELTA editor software.

However, the point of the workflow presented is to provide a simple functionality in an environment well known to most biologist, in order to attract new biologists and educators and increase the production of matrix-based identification tools. Although advanced rules may require some learning effort, it is possible to create useful matrix keys not using these features.

Acknowledgement

This work was supported by the *KeyToNature* Project, ECP-2006-EDU-410019, in the *e*Content*plus* Programme.

References

[1] R. J. Pankhurst, *Practical Taxonomic Computing*, 1991.

[2] J. Winston, *Describing Species*. Columbia University Press,1991.

[3] G. Hagedorn, *Structuring Descriptive Data of Organisms - Requirement Analysis and Information Models*. Ph. D. Thesis, Universität Bayreuth, 2007.

[4] G. Hagedorn, G. Rambold and S. Martellos, "Types of identification keys". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems,* pp. *59-64*, 2010.

[5] G. Hagedorn et al., The Excel to SDD converter, 2010. http://www.keytonature.eu/wiki/Excel_to_SDD_converter, 2010-07.

[6] DELTA – DEscription Language for TAxonomy http://delta-intkey.com/, 2010-07.

[7] G. Hagedorn, G. Weber, A. Plank, M. Giurgiu, A. Homodi, C. Veja, G. Schmidt, P. Mihnev, M. Roujinov, D. Triebel, R. A. Morris, B. Zelazny, E. van Spronsen, P. Schalk, C. Kittl, R. Brandner, S. Martellos and P. L. Nimis, "An online authoring and publishing platform for field guides and identification tools". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 13-18, 2010.

[8] M. Giurgiu, G. Hagedorn and A. Homodi, "IBIS-ID, an Adobe FLEX based identification tool for SDD-encoded multi-access keys". *Proc. of TDWG 2009*, 9-13 Nov. 2009, Montpellier, p. 90, 2009.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 89-93.
ISBN 978-88-8303-295-0. EUT, 2010.

Wiki keys on mobile devices

Gisela Weber, Gregor Hagedorn

Abstract — The development of increasingly powerful mobile devices like PDAs (Personal Digital Assistants) and Smartphones, with larger displays and greater resolution makes them increasingly suitable for identification tools available directly "in the field". One of several approaches towards this aim in the *KeyToNature* project is based on wiki-stored documents. Important features of wiki-based keys, such as hidden text and media information as well as links to glossary entries are supported. The illustrated keys can be used online or downloaded in a zip file. An extension to support the application stores of various mobile platforms (Android, Apple iPhones, etc.) is under development.

Index Terms — mobile devices, smartphones, MediaWiki, identification software Tools.

1 Introduction

Digital identification keys traditionally are built to be used on CD-ROM or over the internet. They have some advantage over printed books in the ease of including a rich selection of colour illustrations, in the speed of updating, and in certain advantages of interactive use, which are especially relevant in pedagogical scenarios. However, they are suitable primarily for indoor workplaces, but usually not practical in the field. With the development of more powerful mobile devices with higher resolution it becomes worthwhile to create digital keys for mobile devices. Such keys must take into account the specific requirements of a small display screen and cumbersome typing to achieve a user-friendly design. In the *KeyToNature* project, several approaches towards this aim have been realized in order to compare different approaches (e.g. [1], [2], [3]).

The application described here is based on the key authoring abilities of the MediaWiki platform described separately [4]. The Wiki is a document storage and authoring platform that allows to embed structured information inside unstructured documents (called "templates"). Based on these structured elements, and taken the freedom of authors to develop new solutions into account, the web-based identification tools are transformed into mobile keys. These can be used online or downloaded packaged into a zip file that can be transferred to mobile devices for offline use.

The authors are with the Julius Kühn-Institute, Federal Research Centre for Cultivated Plants, Institute for Epidemiology and Pathogen Diagnostics, Königin-Luise-Straße 19, D-14195 Berlin. E-mail: gregor.hagedorn@jki.bund.de.

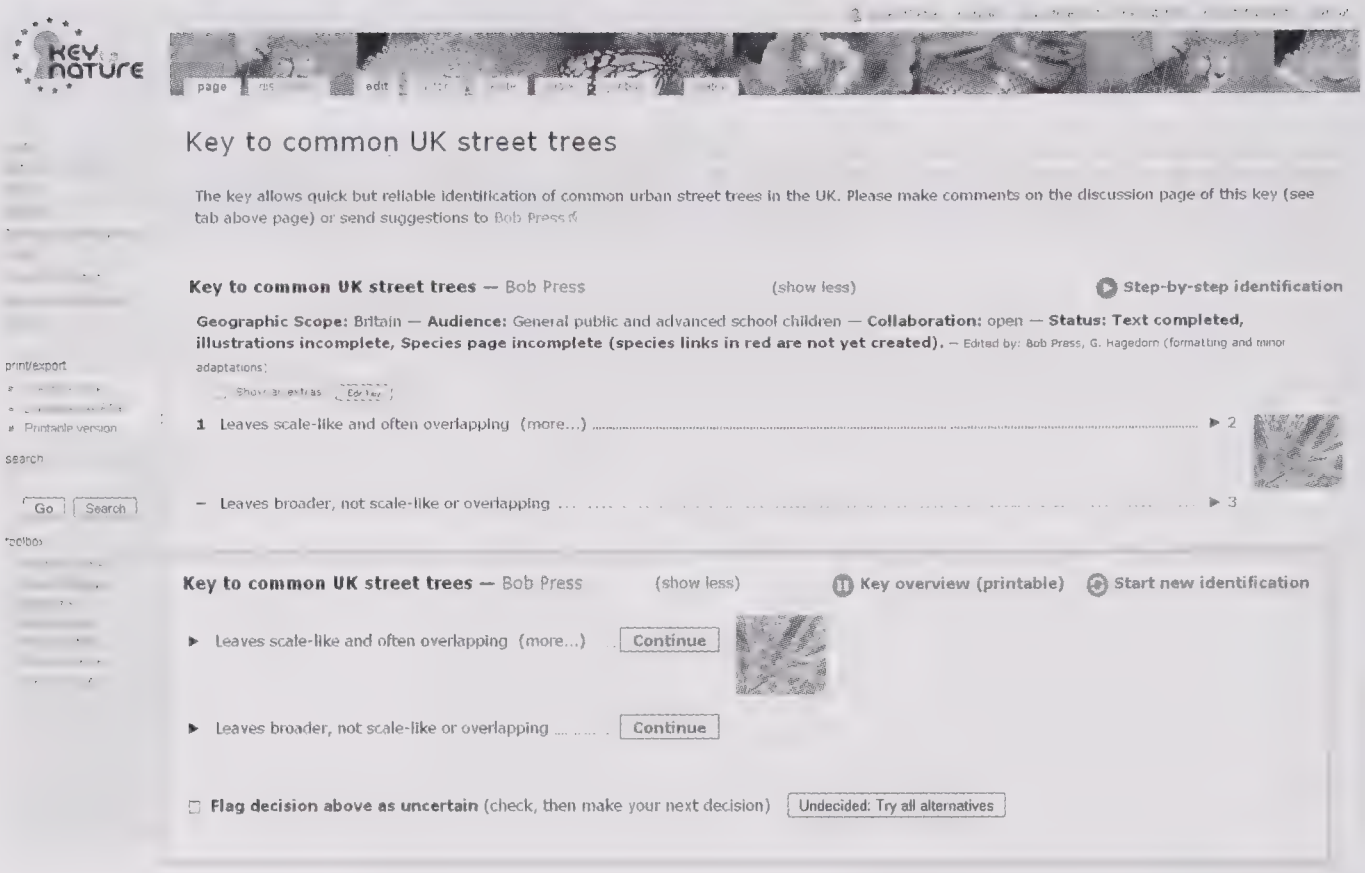

page
edit
Key to common UK street trees
The key allows quick but reliable identification of common urban street trees in the UK. Please make comments on the discussion page of this key (see tab above page) or send suggestions to Bob Press
Key to common UK street trees — Bob Press
(show less)
Step-by-step identification
Geographic Scope: Britain — Audience: General public and advanced school children — Collaboration: open — Status: Text completed, illustrations incomplete, Species page incomplete (species links in red are not yet created). — Edited by: Bob Press, G. Hagedorn (formatting and minor adaptations)
print/export
Printable version
search
Go
Search
1 Leaves scale-like and often overlapping (more...)
2
Leaves broader, not scale-like or overlapping
3
Key to common UK street trees — Bob Press
(show less)
Key overview (printable)
Start new identification
Leaves scale-like and often overlapping (more...)
Continue
Leaves broader, not scale-like or overlapping
Continue
Flag decision above as uncertain (check, then make your next decision)
Undecided: Try all alternatives

Fig. 1 – Wiki key to common UK street trees, printable overview with interactive mode (chosen by upper right “Step-by-step identification” link) shown as overlay in the bottom right.

2 Wiki keys

In the *KeyToNature* project, one approach to enable users to create and edit their own identification keys is the MediaWiki platform. Users can create online single access keys (i.e. dichotomous or polytomous keys) which include images and additional information [5]. These keys can be viewed in a printable overview mode and also interactively in a one-couplet-at-a-time-mode (Fig. 1). Editing of keys occurs online. Special features of the wiki keys are up to 5 images per lead in the right-hand side bar, an extra 2 images below the lead statement, and 6 further images, plus extra text (description, remarks, occurrence) which is initially hidden and requires user interaction to be shown. This principle of showing secondary information only on demand is also used in the display of illustrated term definition (glossary) directly where they are used in the lead statements, and in providing additional information, including legally required IPR and licensing information on the images. All images are zoomable to the maximum possible extent of the source image and display device.

3 Wiki keys on mobile devices

For the transformation of the identification keys from Wiki pages to downloadable HTML-pages optimized for mobile devices we have written a software that is installable as an extension for MediaWiki. This software, called “MobileKey” requires another extension “TemplateParameterIndex”, also written in the *KeyToNature* project, which harvests name-value pairs from template calls on Wiki pages and stores them in a database table. The MobileKey extension uses this index to directly access data behind the Wiki keys. Both extensions of

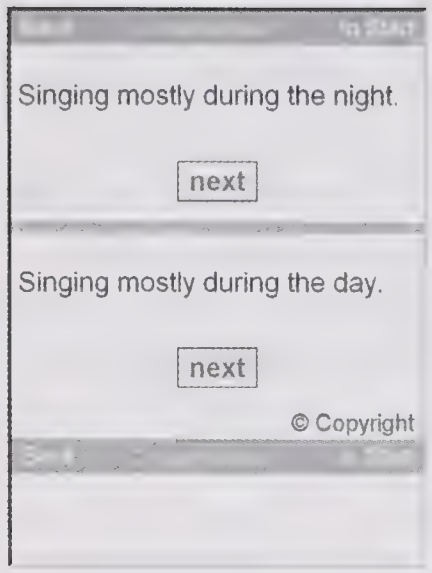


Fig 2 – Couplet with two alternatives (a key to birds based on their sounds).

MediaWiki will be made available as Open Source.

The mobile key extension adds a "Special Page" that allows to create a mobile key that starts at any selected Wiki page with an identification key. The extension:

- recognizes a key on the page and formats the metadata of the key into a start screen
- aggregates the leads that belong to the same couplet;
- splits the key into couplets (= decision), with each couplet being rendered on its own HTML-page in a layout suitable for the small display of mobile devices (Fig. 2);
- puts the additional information on extra HTML pages (Fig. 3, 4);
- puts glossary text on extra HTML pages (Fig. 5);
- stores these files that have been optimized for mobile-devices on the server;
- stores the images on the server;
- replaces the existing links on the Wiki page with the appropriate local links;
- packs all pages and images into a zip file to be downloaded.

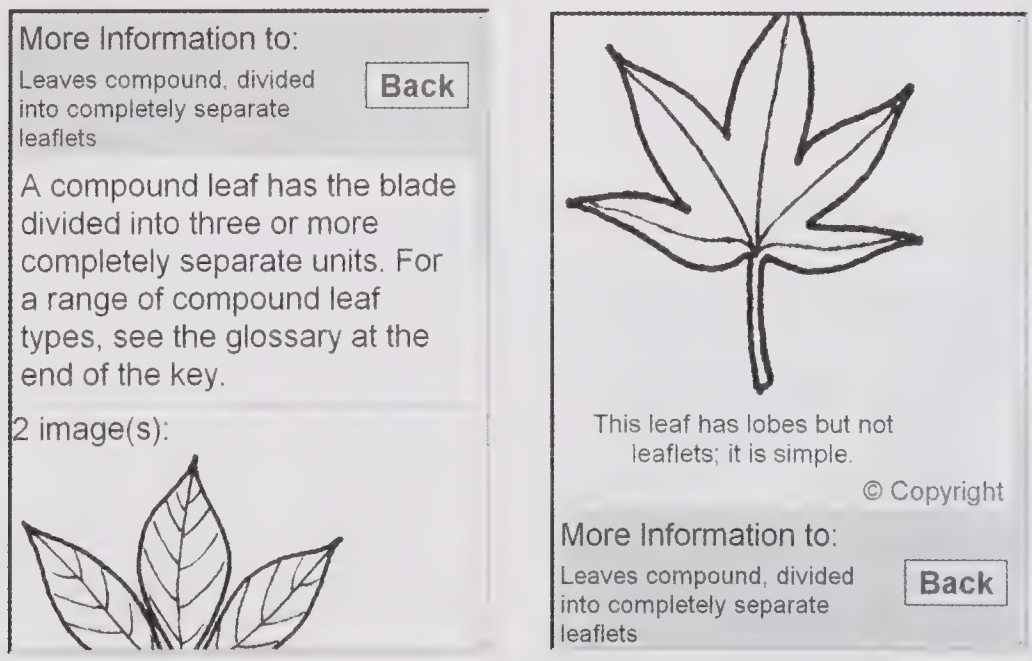

More Information to:
Leaves compound, divided into completely separate leaflets
Back
A compound leaf has the blade divided into three or more completely separate units. For a range of compound leaf types, see the glossary at the end of the key.
2 image(s):
This leaf has lobes but not leaflets; it is simple.
© Copyright
More Information to:
Leaves compound, divided into completely separate leaflets
Back

Fig. 3 – Page with additional information and two images.

The HTML is designed to adjust to some extent to the screen size and landscape versus portrait orientation (Fig. 3, 4). As a mechanism to assess the

display on various devices, the MobileKey Special Page provides two iFrames with different sizes (240 x 320px and 480 x 320px) in which the mobile key can be viewed. Images can be displayed side by side or one below the other according to the display width. Importance is given to good readability of the texts and clear structuring of the displayed page.

The information for a single decision is often longer than the viewport of the mobile device, requiring the user to scroll. At the top of each page, information on how many alternatives are available in the couplet is given. Also the links to go back to the previous couplet or to return from any couplet to the start of the key or a subkey (e.g. for species of a genus) are given there. This information is repeated at the bottom of the page so that the user does not have to scroll all the way up again.

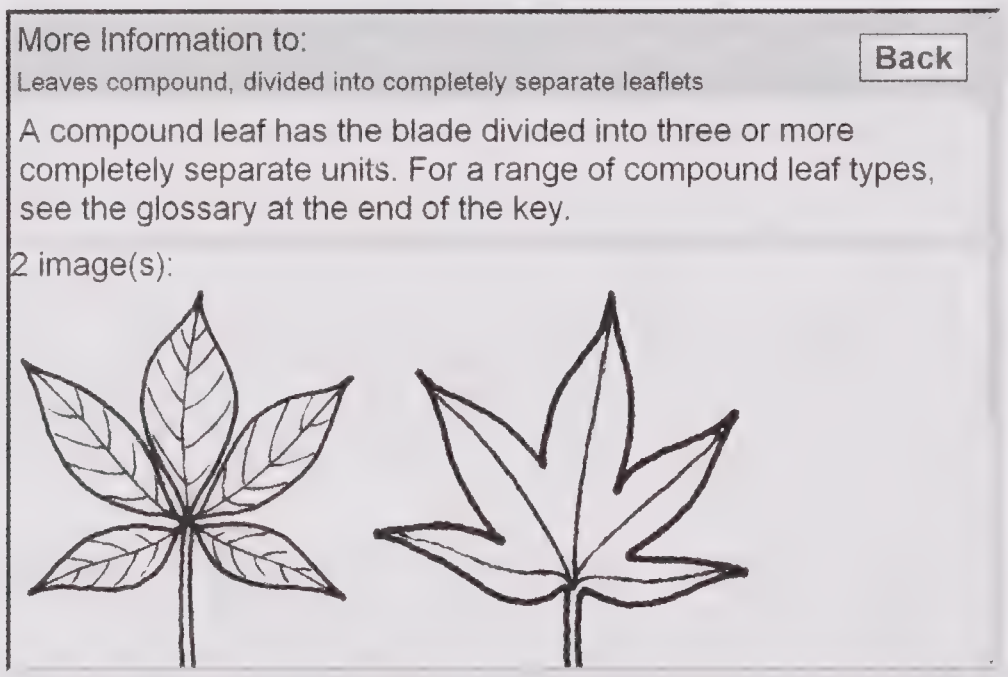

More Information to:
Back
Leaves compound, divided into completely separate leaflets
A compound leaf has the blade divided into three or more completely separate units. For a range of compound leaf types, see the glossary at the end of the key.
2 image(s):

Fig. 4 – Same as Fig. 3, but in landscape orientation.

The bars which contain that information above and below the text are given different colours, supporting the users intuition as to whether a key couplet, a page with additional information or a glossary page is displayed. Only the couplet pages allow navigation within the key or to a subkey, whereas the extra information and glossary pages only offer a link back to the page from which they were called

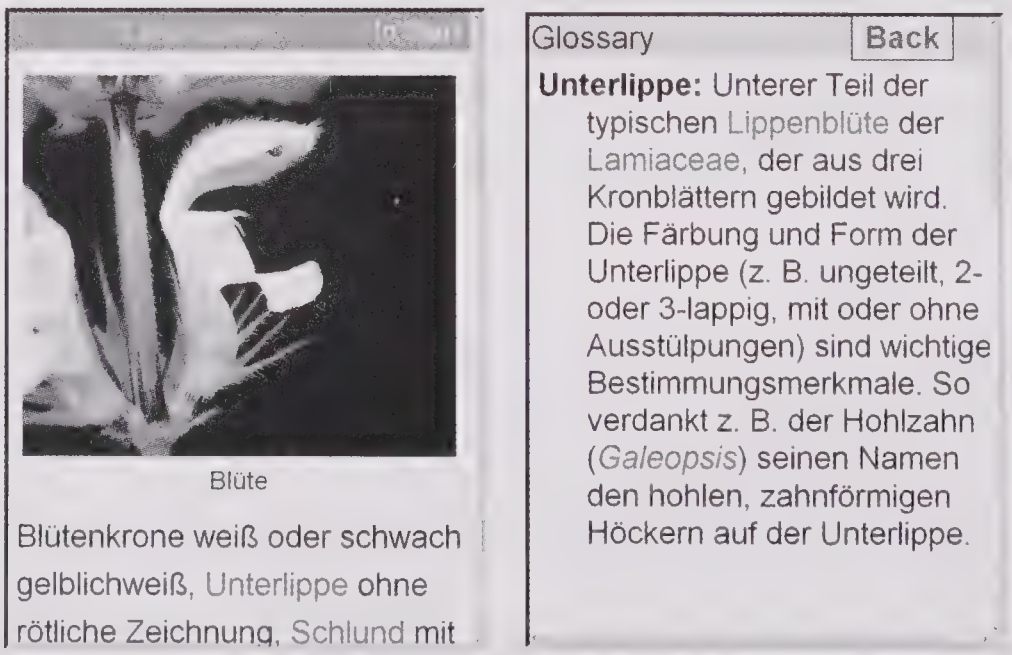

Blüte
Blütenkrone weiß oder schwach gelblichweiß, Unterlippe ohne rötliche Zeichnung, Schlund mit
Glossary
Back
Unterlippe: Unterer Teil der typischen Lippenblüte der Lamiaceae, der aus drei Kronblättern gebildet wird. Die Färbung und Form der Unterlippe (z. B. ungeteilt, 2- oder 3-lappig, mit oder ohne Ausstülpungen) sind wichtige Bestimmungsmerkmale. So verdankt z. B. der Hohlzahn (Galeopsis) seinen Namen den hohlen, zahnförmigen Höckern auf der Unterlippe.

Fig. 5 – Page with glossary links (left, “Unterlippe”, “Schlund”, coloured) and glossary page (right).

4 Outlook

The application for mobile keys is still under development. At the moment, one still has to manually download the zip file to a PC, unzip it, copy the folders to the mobile device's SD card using an USB cable, and manually point the mobile browser to it. It is clear that on mobile devices that support this (especially Android and Apple iPhone), it would be desirable to wrap the identification tools into downloadable mobile apps. In fact, this is the only option iPhones provide.

5 Conclusion

The challenge of developing identification keys for mobile devices is becoming more and more promising with the evolution of better devices. The MediaWiki technology appears to be a good platform to combine user input to create and edit keys with the possibility to make existing keys usable on mobile devices.

Acknowledgement

This work has been supported by the *KeyToNature* project, ECP-2006-EDU-410019, in the frame of the *e*Content*plus* Programme.

References

[1] P. L. Nimis and S. Martellos, "Progetto Dryades", http://www.dryades.eu/home1.html, 2008.

[2] S. Martellos, E. v. Spronsen, D. Seijts, N. Torrescasana Aloy, P. Schalk and P. L. Nimis, "User-Generated Content in the Digital Identification of Organisms: the *KeyToNature* Approach". *International Journal of Information and Operations Management Education (IJIOME)* vol. 3, 3, pp. 272–283, 2010.

[3] E. v. Spronsen, S. Martellos, D. Seijts, P. Schalk and P.L. Nimis, "Modifiable digital identification keys". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems,* pp. 127-131, 2010.

[4] G. Hagedorn, G. Weber, A. Plank, M. Giurgiu, C. Veja, G. Schmidt, P. Mihnev, M. Roujinov, D. Triebel, B. Zelazny, E. v. Spronsen, P. Schalk, C. Kittl, R. Brandner, S. Martellos, P. L. Nimis, "An online authoring and publishing platform for field guides and identification tools". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 13-18, 2010.

[5] G. Hagedorn, B. Press, S. Hetzner, A. Plank, G. Weber, S. v. Mering, S. Martellos, P. L. Nimis, "A MediaWiki implementation of single-access keys". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 77-82, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 95-98.
ISBN 978-88-8303-295-0. EUT, 2010.

A Wiki-based Key to Garden and Village Birds

Tomi Trilar

Abstract — A Wiki-based *Key to Garden and Village Birds* is available in two versions: a dichotomous, hyperlinked and printable version, and as step-by-step identification version. It is supported by *jKey Player* in English, Slovenian, Spanish, Romanian and German.

Index Terms — birds, Aves, key, identification, wiki, jKey Player.

1 Introduction

One of the first steps in discovering and understanding biodiversity is to identify the organisms around us. The study of biodiversity, and in particular the identification of organisms, is becoming part of educational curricula in primary and secondary schools and in many University courses across Europe.

After Gutenberg, information useful for identifying organisms was printed on paper. The constraints of a paper-printed text have forced most authors to organise information according to the hierarchical scheme of biological classification. Computer-assisted keys, on the contrary, allow us to identify organisms without necessarily using the characters of the classification in the biological system. Classification and identification belong to two different operational processes. Classification is the job of taxonomists, identification can be fun for anybody [1].

There are many identification tools available online, based on different platforms. One of the possible platforms is the Wiki. The advantage of wiki-based keys is that a community of users can work together in constructing, improving and enriching them in a collaborative way.

In the European project *KeyToNature,* two tools for creating and running wiki-based keys were developed: the *jKey Editor* and the *jKey Player*. The JavaScript based *jKey Player* is supporting interactive use, tracking and annotating previous decisions, and allows all decision steps to be easily revisable. The revision capability is especially useful in classroom use, allowing the quick identification of erroneous decisions, discussing the reasons for these, and continuing without

The author is with the Slovenian Museum of Natural History, Prešernova 20, P.O.Box 290, 1001 Ljubljana, Slovenia. E-mail: ttrilar@pms-lj.si.

restarting the entire identification process. Additional functionalities like toggling the display of secondary text and images, and enlarging images on click, make the keys more user-friendly. The complementary web-based editing tool *jKey Editor* allows a form-based editing of the identification keys, which simplifies creating, modifying, pedagogically adapting, or translating existing keys [2].

2 Key to Garden and Village Birds

The *Key to Garden and Village Birds* is a wiki-based identification tool, which includes text, pictures, drawings and sounds [4]. Original recordings of bird sounds are also displayed, which were stored in the Slovenian Wildlife Sound Archive, housed in the Slovenian Museum of Natural History (PMSL) in Ljubljana, Slovenia [3].

The key can be used both in a dichotomous hyperlinked printable version (Fig. 1) or in a step-by-step identification version (Fig. 2) supported by *jKey Player*. The target audience are pupils in primary and secondary schools, with the pedagogical goal to observe the birds and listen their singing and calling. Important was the contribution of the wiki-community, who translated the key into English [4], Slovenian [5], Spanish [6], Romanian [7] and German [8].

page · discussion · view source · history

Key to Garden and Village Birds

Key to Common Garden and Village Birds of Europe — Tomi Trilar

Step-by-step identification

(show less)

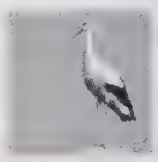
This key is limited to observations of birds that are singing or otherwise producing sounds with their bills.
Geographic Scope: Central Europe — **Source:** (original work) — **Collaboration:** open — **Status: Under testing: Sounds of 3 species are missing. Species pages not done (species links in red = not yet created).** — Edited by: Tomi Trilar, Gregor Hagedorn (formatting and minor adaptations), Bob Press (style and language improvements)

Show all extras

1 a Bird either clattering its bill or drumming on wood. ..... ☐ 2

1 b Bird singing or calling. ..... ☐ 5

2 a Black and white bird, 100 to 115 cm tall, with half of that height being made up by the red legs. Often clattering its bill. ..... **White Stork (*Ciconia ciconia*)**



2 b Bird less than half as large, usually climbing the trees using stiff tail as a support. Drumming rhythmically on tree trunks (woodpeckers). ..... ☐ 3

3 a Back black with the white spots. Drums in spring, the drum rolls characteristically short. Song consists of short, sharp, high-pitched notes. ..... **Great Spotted Woodpecker (*Dendrocopos major*)**

3 b Back green. ..... ☐ 4

Fig. 1 – Dichotomous hyperlinked printable version of the wiki-based *Key to Garden and Village Birds* [4].

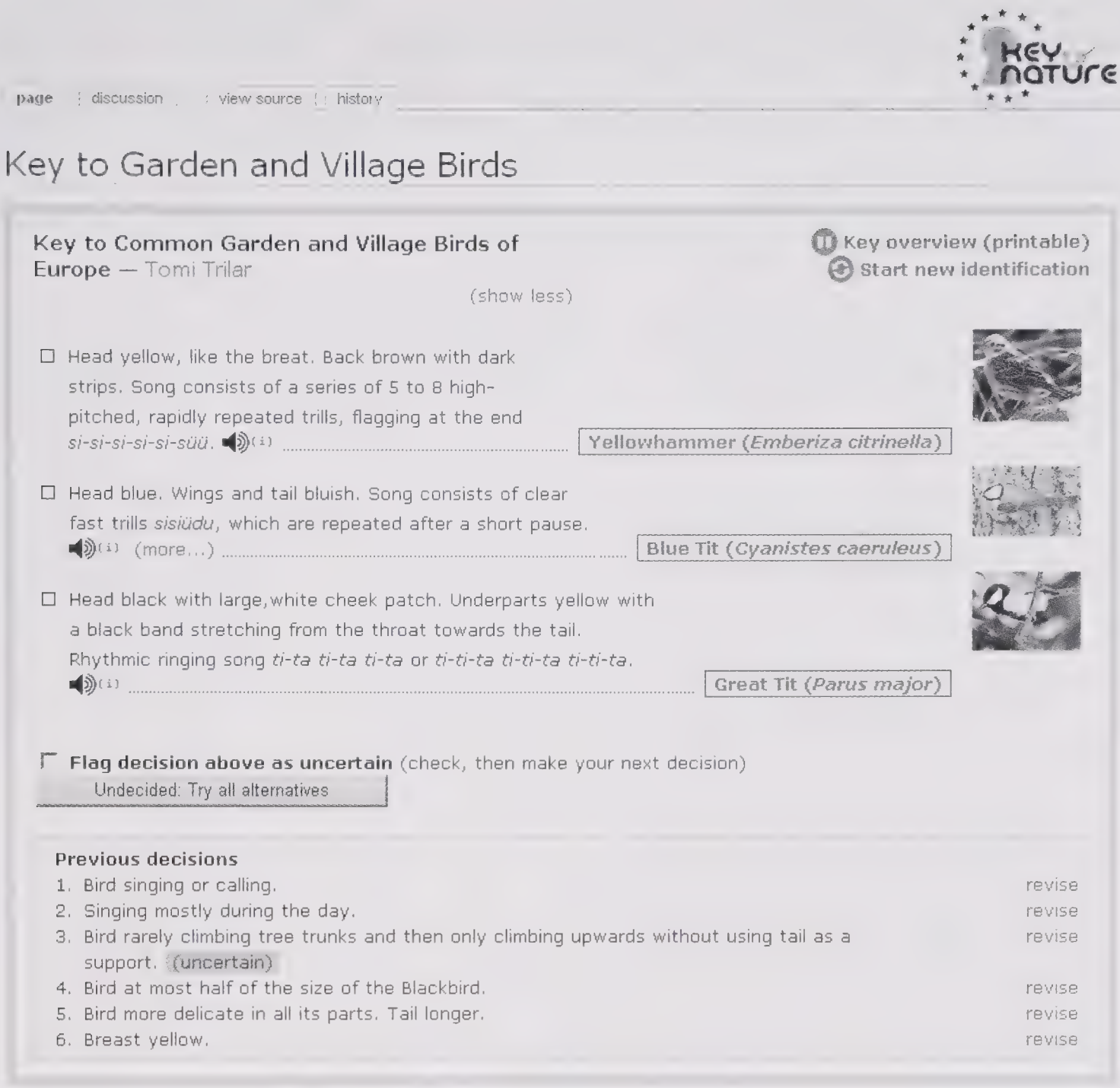

page
discussion
view source
history
Key to Garden and Village Birds
Key to Common Garden and Village Birds of Europe — Tomi Trilar
(show less)
Key overview (printable)
Start new identification
Head yellow, like the breat. Back brown with dark strips. Song consists of a series of 5 to 8 high-pitched, rapidly repeated trills, flagging at the end si-si-si-si-si-süü.
Yellowhammer (Emberiza citrinella)
Head blue. Wings and tail bluish. Song consists of clear fast trills sisiüdu, which are repeated after a short pause.
(more...)
Blue Tit (Cyanistes caeruleus)
Head black with large,white cheek patch. Underparts yellow with a black band stretching from the throat towards the tail.
Rhythmic ringing song ti-ta ti-ta ti-ta or ti-ti-ta ti-ti-ta ti-ti-ta.
Great Tit (Parus major)
Flag decision above as uncertain (check, then make your next decision)
Undecided: Try all alternatives
Previous decisions
1. Bird singing or calling. revise
2. Singing mostly during the day. revise
3. Bird rarely climbing tree trunks and then only climbing upwards without using tail as a support. (uncertain) revise
4. Bird at most half of the size of the Blackbird. revise
5. Bird more delicate in all its parts. Tail longer. revise
6. Breast yellow. revise

Fig. 2 – Step-by-step identification version of the wiki-based *Key to Garden and Village Birds* with an annotated step of uncertainty [4].

Acknowledgement

I would like to thank Gregor Hagedorn for encouraging me to create the key and for his help in formatting and dapting it, and Bob Press for style and language improvements. I would like to also thank Marina Ferrer Canal, Mircea Giurgiu, Gregor Hagedorn and Irena Kodele Krašna, who organised and helped with the translation of the key into Spanish, Romanian, Slovenian and German. This work was supported by the *KeyToNature* Project, ECP-2006-EDU-410019, in the *e*Content*plus* Programme.

References

[1] P. L. Nimis, *Keys to the Lichens of Italy*. Ed. Goliardiche, Trieste, 341 pp., 2004.

[2] G. Hagedorn and S. Opitz, "JKey Player", *KeyToNature*. http://www.keytonature.eu/wiki/JKey_Player, 2010.

[3] "Slovenian Wildlife Sound Archive", Slovenian Museum of Natural History. http://www2.pms-lj.si/staff/bioacoustics/bioacoustics.html, 2010.

[4] "Key to Garden and Village Birds", *KeyToNature*. http://www.keytonature.eu/wiki/Key_to_Garden_and_Village_Birds, 2010.

[5] "Ključ za določanje vrtnih ptic", *KeyToNature*. http://www.keytonature.eu/wiki/Ključ_za_določanje_vrtnih_ptic, 2010.

[6] "Clave de Aves Comunes de Jardines y Areas Rurales de España", *KeyToNature*. http://www.keytonature.eu/wiki/Clave_de_Aves_Comunes_de_Jardines_y_ Areas_Rurales_de_España, 2010.
[7] "Cheie de identificare a pasarilor comune care traiesc in zone rurale din Europa", *KeyToNature*. http://www.keytonature.eu/wiki/Cheie_de_identificare_a_ pasarilor_comune_care_traiesc_in_zone_rurale_din_Europa, 2010.
[8] "Häufige Vögel in Gärten und Siedlungen", Offene-Naturführer. http://www.offene-naturfuehrer.de/wiki/Häufige_Vögel_in_Gärten_und_Siedlungen, 2010

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 99-105.
ISBN 978-88-8303-295-0. EUT, 2010.

Wiki-keys for the ferns of the Flora of Equatorial Guinea

Francisco Cabezas, Carlos Aedo, Patricia Barberá,
Manuel De la Estrella, Maximiliano Fero, Mauricio Velayos

Abstract — Flora of Equatorial Guinea is a research project coordinated by the Real Jardín Botánico of Madrid with the aim to produce a modern Flora for this almost unknown territory. One of our goals is to develop the website www.floradeguinea.com. A system of relational databases with an implemented Thesaurus - to minimize typographical mistakes - allows the on-line managing and updating of nomenclatural information, herbarium specimens, literature records, vernacular names and bibliography. Images and maps are also linked to accepted names. We are now implementing interactive keys for identification. In our wiki-based system, Scans of herbarium specimens from all species growing in the country are also being uploaded. Species from neighboring countries are included as well, since some of them could appear in Equatorial Guinea.

Index Terms — Africa, Equatorial Guinea, e-Flora, ferns, online-databases, Wiki key.

1 Introduction

Flora of Equatorial Guinea is a research project undertaken by the *Real Jardín Botánico de Madrid*-CSIC, with the collaboration of the Spanish Universities of Salamanca and Córdoba, the Kew Royal Botanic Gardens, the *Nationaal Herbarium Nederland* and the *Université Libre de Bruxelles.* The project is currently funded by the Spanish administration (reference project CGL2009-07405). The final aim of our project is to produce a modern flora of this almost unknown territory, a historic goal of the Spanish botany, rooted in former times, when these tropical regions were part of the overseas provinces of Spain, more than sixty years ago.

The interest on this region also derives from its important biodiversity. Tab. 1 shows the surface of humid rain-forests in some countries of Central Africa: only Gabon has a higher percentage of the territory covered by humid, undisturbed

F. Cabezas is with the Department of Botany, Faculty of Biology, University of Salamanca. Méndez Nieto Av. s/n, 37007 Salamanca, Spain. E-mail: fcabezas@usal.es.
C. Aedo, P. Barberá, M. Fero and M. Velayos are with the Real Jardín Botánico de Madrid, CSIC. Murillo Sq. 2 28014, Madrid, Spain.
M. Estrella is with the Botany, Ecology and Plant Physiology Department, C-4, Celestino Mutis, Campus de Rabanales, 14071, Córdoba, Spain.

rain-forests in the Guineo-Congolian region, the most biodiverse area in the mainland of tropical Africa.

	Original tropical rainforest surface (km^2)	Tropical rainforest surface nowdays (km^2)	% of tropical rainforest preserved
Cameroon	376.900	155.330	41,2
Equatorial Guinea	26.000	17.004	65,4
Central African Republic	324.500	52.236	16,1
Rwanda	9.400	1.554	16,5
Nigeria	421.000	38.620	9,2
Gabon	258.000	227.500	88,2

Tab. 1 – Surface of humid rain-forest in some countries of Central Africa.

Our interest was also increased by the dissimilar floristic knowledge of the different regions of the country, reflecting once more the complicated history of the territories of Equatorial Guinea and Spain since Emilio Guinea's first trip, including questions beyond science, for example, such as the independence of Equatorial Guinea from Spain in 1968.

One of the goals of the planned modern Flora of Equatorial Guinea is the development of our website, www.floradeguinea.com, where new identifications are updated immediately. Any specialist working on the flora of Africa can freely check our results.

Presently, we are determined to go one step forward, implementing in our site a on-line interactive system with identification keys. In our wiki-based system, scans of herbarium specimens from all species growing in the country are also uploaded. Species from Gabon, Cameroon and S. Tomé & Príncipe are included as well: some of them could appear in Equatorial Guinea, although they have not been collected yet.

2 Materials and methods

We have planned our task in 4 different steps: 1) compilation of all literature records, 2) identification of all the available specimens, 3) compilation of checklists of taxa and specimens growing in the country, 4) compilation of a Flora, with keys for identification, descriptions, images and maps.

In order to reach the first two goals, we have designed a system of relational databases with several Thesauri, implemented to minimize typographical mistakes. Nomenclatural information, specimens data, literature records, vernacular names and bibliography can be managed and updated on-line. Digital images and maps with the distribution in Equatorial Guinea are also

linked to accepted names.

The second two goals (checklist and Flora) demanded a re-designing of the whole database structure. We used a wiki-template designed by Gregor Hagedorn and used by *KeyToNature* and other projects in the Real Jardín Botánico-CSIC, in order to translate our printed keys into an interactive system where updates done by users or editors can be automatically implemented. This also permits to include more information in the e-keys, such as e.g. scans of herbarium specimens.

The final output of the keys will be linked to the www.floradeguinea.com site, where, following the taxon name, the user can find revised nomenclatural information, a description, a list of identified herbarium specimens, a list of literature reports, digital images of herbarium specimens, links to pictures of living plants, and a skecth map where the distribution of the taxon in Equatorial Guinea is presented.

3 Results and Discussion

3.1 Results of the Flora of Equatorial Guinea

The gathering of literature produced a great stepping stone with the publication of the documentary databases for the Flora of Equatorial Guinea [1]. In this book, all the reports of mosses, fungi, ferns, mono- and dicotyledons from Equatorial Guinea were compiled and databased, reflecting the information as it was originally published. Widely distributed taxa from neighboring countries were also included, since they will appear in the keys for identification. Currently, this database is still growing and is managed completely on-line, since it needs to be updated with any new publication on floristics for any of the territories of Equatorial Guinea. We also go on compiling and including the names published in São Tomé and Príncipe Island, Cameroon, and Gabon. Today, 52,301 records of vascular plants are included.

The second task (collection effort), has produced as a main result a set of more than 15,000 collection numbers from Equatorial Guinea in the Madrid herbarium, with an average of 4 duplicates: this is now the main collection for the country. In our database we also include the vouchers data, especially from historical collections kept at K, BM, or in the Netherlands in the WAG herbarium, as well as those kept in Equatorial Guinea. For comparative purposes, some collections from neighboring countries were also studied. In this aspect, Missouri and Portuguese herbaria were essential, and are now databased as well. Until now, 16,615 herbarium specimens from Equatorial Guinea have been databased. Among them, about 2000 were studied carefully and assigned to a correct name.

The lack of floristic knowledge in Africa and, of course, in Equatorial Guinea, brought us to publish critical checklists before the Flora, splitting the original idea of Emilio Guinea. The main reason is clear: if we'll have waited until the information is complete to publish the Flora, and the identification keys and descriptions are made, most of the species included in the work could be extinct.

These checklists, on the contrary, provide a useful tool to start with conservation programs and strategies. The latter are especially necessary nowadays, considering that 13 million ha of primary forests are destroyed every year. This step was the one with more results in the last years.

	FWTA	Fl. GUI	Increase (%)	Nrec Country
Pteridophytes	163	226	38	8
Cyperaceae	28	96	231	22
Marantaceae	4	26	271	8
Piperaceae	9	13	44	_
Mimosoideae	9	40	344	14
Caesalpinioideae	23	124	525	45
Papilionoideae	121	157	30	48
Ebenaceae	1	28	2700	12
Melastomataceae	18	57	216	26
Commelinaceae	24	45	114	11

Tab. 2 – Numerical summary of families with checklist published since 2001. FWTA= Number of species mentioned in Equatorial Guinea in the *Flora of West Tropical Africa*. Fl.GUI=Number of species found in our study. Increase = percentage increase of our catalogue compared with the data of FWTA. Nrec Country = Number of species found in our catalogues not previously reported from Equatorial Guinea. Families in boldface include new names or species described from material collected in Equatorial Guinea.

Tab. 2 shows the numerical summary of families where checklists were published since 2001. We have published the accounts of 24 families of vascular plants and all Pteridophytes. Our main result was that floristic knowledge of the country increased by about 120 %. Applying this value to the estimation of Davis [2] of 3,250 species growing in Equatorial Guinea, using data relative to the Flora of West tropical Africa, the number of species estimated to live in the country jumps to c. 7,100, a figure similar to those of Cameroon and Gabon.

The final task is that of writing and editing the Flora. Until now, we have published two volumes of the flora, Pteridophytes and Fabaceae, with 227 and 320 species, respectively. These figures include also citations of species without herbarium support, whose distribution range and habitat render their presence in Equatorial Guinea quite possible. Moreover, several monographic treatments are being undertaken on some genera, which will probably result in the reinterpretation of several taxa. Once the first reference is published, it is easier to compare any new collection, so that we would be not surprised if these numbers will be changing soon.

3.2 Results in the website

The development of a new system of databases, based on MySQL allows the on-line handling of information related to herbarium specimens, literature reports, and also of those data related to accepted names, bibliography, type information or descriptions. We have included in the website all the information that we have produced in the last years, so that now all papers produced by us or by other authors but related to the country are freely downloadable [3-18].

The next step is to produce and update all the stages of the editorial process on-line. We have implemented the printed version of the keys, and now we are presenting interactive keys as well [http://www.keytonature.eu/wiki/Clave_de_familias_de_Pteridophyta_de_la_Flora_de_Guinea_Ecuatorial_%28RJB%29]. These keys are based on the wiki-structure design developed by Gregor Hagedorn within the European project *KeyToNature*, which have been used also by other projects in the *Real Jardín Botánico*, as the Gymnosperms key ["*Clave de Gimnospermas* (RJB)"].

The printed keys of the published volumes of the Flora are being translated to this wiki system in a relatively easy way. Until now, we have migrated all the keys for identification of ferns. In this first step the keys are in Spanish, the mother language in Equatorial Guinea, but English versions will be also prepared. We have material for starting with the translation of the key for families of vascular plants and of all the keys produced for the volume of legumes. At the end all the keys included in the Flora of Equatorial Guinea will be produced, updated and corrected on-line, and in the future from the on-line version a printed version will be published. This system allows to advance the complicated editorial process of a work like a Flora without the constraints of printed texts.

In our keys, all species growing in Equatorial Guinea are included, and a digital picture of an herbarium specimen is included close to each name. Each result in the key (family, genus or species name) is linked to a taxon page at the www.floradeguinea.com site, which has information related to nomenclature, specimens, literature records and bibliography or vernacular names. Furthermore, each accepted name is linked to several sites where more information can be consulted, such as the African Plants Database, by the *Conservatoire et Jardin Botanique of Geneva* (http://www.ville-ge.ch/musinfo/bd/cjb/africa/recherche.php), where accepted names for the most important african floras are included and critically presented, or the West African Plants Database developed by the Senckenberg Forschungsinstitut und Naturmuseum and by the Frankfurt University (http://www.westafricanplants.senckenberg.de/root/index.php), where, among other information, pictures of living plants can be examined.

4 Conclusions

High speed degradation of natural resources and the need of urgent conservation decisions have increased the value of Floras as the base to study, understand and preserve plant biodiversity.

Despite their high relevance, many Floras remain incomplete and progress

slowly, due to the large number of species involved, the highly distributed and dissimilar data, and the lack of tools promoting effective collaboration. Thus, is common to find both personal and duplicated efforts between remote researchers.

Nowdays, in order to make the information contained in old floristic works accesible, Floras only available as hard copies are increasingly being digitized. Most of the initiatives have produced scans or fixed images of the original printed version. With the use of new technologies, printed versions are not the unique result of a flora, they appear to be just one of the possibilities.

This can be overcome by making full use of current information technology to draw the highly distributed data together and to allow the taxonomic research community to communicate efficiently. The development of an e-way of handling Floras will change traditional work-flow processes by fostering a collaborative setting, strengthening existing research networks, and making plant biodiversity information rapidly and widely accessible in a re-usable format.

Floras provide keys for the identification of plant species and additional information on each species such as synonymy, economic uses, geographical distribution and ecology. With the possibility of e-handling, floristic research will be optimized by the producers of Floras, and the consumers of information will increase. Potential users will range from traditional readership to ecologists interested in morphological traits, climate modellers or policy makers interested in distribution data.

The benefits of the wiki-keys will be:

1. More efficient production of floras, especially the keys for identification. Keys will be more interactive and easily updated, making a more dynamic and updated floristic resource in a collaborative system.

2. The output information of the work is higher and easier to access. The impact of floristic data will be greatly increased. This is providing novel uses of floristic data. The availability of digital floristic data opens opportunities for uses as i.e. datasets for modelling and web services connecting to other websites and databases.

3. The wiki-key system also promotes new collaboration mechanisms for taxonomists in order to significantly improve local and remote co-working and to eliminate redundancy of work within this scattered community. Collaborative research will foster the transfering of knowledge to the benefit of early-stage researchers, particularly in tropical and developing countries.

Acknowledgement

The authors wish to thank to the Ministry of Science and Innovation of Spain for the support in the next years. We are also indebted to the Spanish Superior Research Council, especially the Department of Publications. The authorities of Equatorial Guinea, the people responsible of BATA herbarium and the National University of Equatorial Guinea (UNGE) deserve special mention. The vouchers scanned and used are mainly from our collection in MA. Nevertheless, as can be inferred from some headings, some species were obtained from abroad institutions as: Botanischer Garten und Botanisches Museum of Berlín (B), Botanische Staatssammlung München (M), Université Libre de Bruxelles (BRLU), Royal Botanic Gardens, Kew (K), Natural History Museum of London

(BM), National Botanic Garden of Belgium (BR), Muséum National d'Histoire Naturelle, Paris (P), Wageningen University (WAG) y Cameroon National Herbarium, Yaounde (YA). All of them are thanked for the permissions to use their herbarium scans.

References

[1] C. Aedo, M. Velayos and M. T. Tellería (eds.), *Bases documentales para la Flora de Guinea Ecuatorial. Plantas vasculares y hongos. Madrid*, Consejo Superior de Investigaciones Científicas & Agencia Española de Cooperación Internacional, 414 pp., 1999.

[2] S. D. Davis, V. H. Heywood and A. C. Hamilton (eds.), *Centres of Plant Diversity, A Guide and Strategy for their Conservation, Europe, Africa, South West Asia and the Middle East*. Cambridge: World Wide Fund for Nature (WWF) and The World Conservation Union (IUCN), vol. 1, 1994.

[3] M. Fero, F. Cabezas, C. Aedo and M. Velayos, "Checklist of the Piperaceae of Equatorial Guinea", *Anales Jard. Bot. Madrid*, vol. 60(1), pp. 45-60, 2003.

[4] I. Parmentier and D. Geerinck, "Checklist of the Melastomataceae of Equatorial Guinea", *Anales Jard. Bot. Madrid,* vol. 60(2), pp. 331-346, 2003.

[5] F. Cabezas, C. Aedo and M. Velayos, "Checklist of the Cyperaceae of Equatorial Guinea (Annobón, Bioko and Río Muni), *Belg. J. Bot.*, vol. 137(1), pp. 3-26, 2004.

[6] F. Cabezas, M. Estrella, C. Aedo and M. Velayos, "Marantaceae of Equatorial Guinea", *Ann. Bot. Fennici*, vol. 42(3), pp. 173-184, 2005.

[7] B. Senterre, "Checklist of the Ebenaceae of Equatorial Guinea", *Anales Jard. Bot. Madrid,* vol. 62(1), pp. 53-63, 2005.

[8] M. Estrella, F. Cabezas, C. Aedo and M. Velayos, "Checklist of the Mimosoideae of Equatorial Guinea", *Belg. J. Bot*, vol. 138(1), pp. 11-23, 2005.

[9] M. Estrella, F. Cabezas, C. Aedo and M. Velayos, "Checklist of the Caesalpinioideae (Leguminosae) of Equatorial Guinea (Annobón, Bioko and Río Muni)", *Bot. J. Linn. Soc., v*ol. 151, pp. 541-562, 2006.

[10] A. P. Davis and E. Figueiredo, "A checklist of the Rubiaceae (coffee family) of Bioko and Annobon (Equatorial Guinea, Gulf of Guinea)", *Syst. Biodivers.,* vol. 5(2), pp. 159-186, 2007.

[11] F. Cabezas, M. Estrella, C. Aedo and M. Velayos, "Checklist of the Commelinaceae of Equatorial Guinea (Annobón, Bioko and Río Muni)", *Bot. J. Linn. Soc.,* vol. 159, pp. 106-122, 2009.

[12] P. Jiménez-Mejías and F. Cabezas, "*Schoenoplectus heptangularis* Cabezas & Jiménez Mejías (Cyperaceae), a new species from Equatorial Guinea", *Candollea,* vol. 64, pp. 101-115, 2009.

[13] M. Estrella, F. Cabezas, C. Aedo and M. Velayos, "The Papilionoideae (Leguminosae) of Equatorial Guinea (Annobón, Bioko and Río Muni)", *Folia Geobot.*, vol. 45, pp. 1-57, 2010.

[14] M. E. Leal, "Novitates Rio Munis 1. A new endemic *Scaphopetalum* (Malvaceae) from Mount Mitra, Equatorial Guinea", *Blumea*, vol. 52, pp. 137-138, 2007.

[15] E. Figueiredo, A. Gascoigne and J. P. Roux, "New records of Pteridophytes from Annobón Island", *Bothalia*, vol. 39,2, pp. 213-216, 2009.

[16] M. S. M. Sosef and N.S. Nguema Miyono, "Novitates Rio Munis 2. A new species of *Begonia* section *Loasibegonia* (Begoniaceae) from the Monte Alen region, Equatorial Guinea", *Blumea*, vol. 55, pp. 91-93, 2010.

[17] M. Estrella, C. Aedo, B. Mackinder and M. Velayos, "Taxonomic Revision of *Daniellia* (Leguminosae: Caesalpinioideae)", *Syst. Bot.,* vol. 35(2), pp. 296-324, 2010.

[18] T. Stévart, V. Cawoy, T. Damen and V. Droissart, "Taxonomy of Atlantic Central African Orchids 1. A New Species of *Angraecum* sect. *Pectinaria* (Orchidaceae) from Gabon and Equatorial Guinea", *Syst. Bot.*, vol. 35(2), pp. 252-256, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 107-112.
ISBN 978-88-8303-295-0. EUT, 2010.

MyKey: a server-side software to create customized decision trees

David Gérard, Régine Vignes Lebbe

Abstract — To facilitate the identification of specimens, biodiversity informatics has developed numerous new computer-aided tools which compete with the old, printed single access keys. Free access keys are accessible for many different taxonomic groups, but key-generating software are also helpful to construct single access keys. This paper presents a midfield solution to create customized decision trees (single access keys) through an interactive web-based interface. This solution offers an online service to create keys according to the parameters and context chosen by the final users themselves. It is also useful for the administrator of this online system, because of low-maintenance needs, limited to configuration files when new knowledge bases are added on the server side. Presently, the software is available with a French interface at the following URL: http://baron.snv.jussieu.fr/cgi-bin/david/MyKey.cgi.

Index Terms — polytomous key, customized key, decision tree, web-based interface.

1 Introduction

Accessing relevant and critical taxonomic information is often a privilege for the specialists [1], who can take profit of natural history collections, taxonomic monographs or low-circulation journals. Lawyers, border guards, epidemiologists, as well as ecologists or any other biologists, may also have identification requirements. For their needs, printed dichotomous or polytomous keys are often included in monographs, floras and faunas, and in practical field guides. A key has a graph structure, comparable to a decision tree of Artificial Intelligence [2] (in this paper we will use as synonyms the terms single access key, and decision tree).

A negative property of classical keys is their static nature: if you cannot answer to a question (for example if you have no flower and characters of the flowers are frequently used in botanical keys), the key is useless. Moreover, to create a

D. Gérard was student in the Laboratoire Informatique et Systématique, University of Paris 6, UMR7207 (MNHN, CNRS, UPMC), CP 48, 57 rue Cuvier 75005 Paris, France. E-mail: dagerard@gmail.com.
R. Vignes Lebbe is with the University of Paris 6, UMR7207 (MNHN, CNRS, UPMC), CP 48, 57 rue Cuvier 75005 Paris, France. E-mail: regine.vignes_lebbe@upmc.fr.

key is a time-consuming task and each taxonomist adapts his/her key to a given context. But it could be necessary to offer different keys to different user groups (e.g. an autumn key to trees based only on trunk and leaves characters, a key based on fruits, a key limited to a geographic area, a key taking into account immature stages, etc.).

In the late 1960s, biologists [3], [4] began to use computers to produce more flexible free access keys or computer-aided-identification (CAI) systems [5]. Since the 1980s, knowledge bases formats for structuring descriptive data appeared, like the DELTA format [6], and user-friendly software (e.g. IntKey [7] and XPER [8]) were implemented for creating knowledge bases and enabling CAI. Storage of data became easier, as well as the retrieving of specific information in a pool of data [9]. The reader can find a good report comparing these tools in [10], [11], on the DELTA website, and on the BD tracker of the European project EDIT [http://www.e-taxonomy.eu/].-

Causse & Lebbe [12] have demonstrated the strong similarity between CAI and single access keys, and their common elimination procedure. These authors introduce the idea of a unique system able to propose identification from a free access key to a single access key by continuously improving the strategy advertisement expressed by the taxonomist.

To adapt single access keys to the users, one finds different proposals (see for example [18] and [19] in this book). This paper offers another solution: it combines a program to compute automatically single access keys and a web interface for the final user to define himself the input parameters of the key constructor. This original prototype, MyKey, is a server-based program. It uses knowledge bases stored with the XPER system and the key generator MAKEY [13]. Running on a server of the Laboratoire d'Informatique et Systématique of the University Paris 6, it is available at the following URL (http://baron.snv.jussieu.fr/cgi-bin/david/MyKey.cgi).

2 The tools XPER and MAKEY

XPER and its current version Xper² 2.1 is a complete software package for managing knowledge bases [14]. It provides tools for structuring and using taxonomic descriptions and for identifying specimens. The basic program allows to save structured descriptions, comparable with the DELTA format and consisting of three main elements: taxa, descriptors (characters for DELTA) and states of descriptors. An import/export in SDD XML schema the new standard proposition of TDWG, is also available (SDD= Structured Descriptive Data. See : http://wiki.tdwg.org/SDD).

MAKEY [13], [15] is a key generator software. It selects step by step the best question (descriptor) to create a node. By default, the choice is based on the discriminating power (ability to split the pool of taxa in equal disjointed classes, two or more classes) and so MAKEY tries to generate short and well-balanced polytomous keys [16].

MAKEY creates keys discriminating the different taxa of the input knowledge base. Other versions create a key to discriminate groups or classes of the taxa; these classes are sets of taxa defined by different character states of a selected

descriptor. For example, if a descriptor is the toxicity of mushrooms, MAKEY can create a key to identify the toxicity of a mushroom even if the specimen is not identified at the species level; in the same manner it is possible to create keys to identify genera within a knowledge base describing species if a descriptor associates each species to its genus. The manual to use MAKEY is accessible on line.

3 Description of MyKey

The final user of a key is the best person to know his observation constraints. So, the concept of the MyKey service was to offer to the final user the possibility to create identification keys customized to his needs.

A web interface gives an access to the different input parameters of the Makey software. We classify these parameters in four categories:

- parameters are related to the data coverage of the key; for example one can generate a key to all species of the knowledge base or a key restricted to a given geographical area or to a genus etc.
- parameters are related to the taxonomic domain (importance and easiness to observe a character),
- parameters have consequences on the topology of the key, like the criterion to select characters (by default it minimizes the mean number of questions to achieve an identification),
- parameters concern the format of the result: indentented or bracketed key (see [21] in this book), text or HTML format etc.

So the interface is divided in four parts according to these categories.

3.1 Goal or terminal nodes of the key

The user selects the goal of the discrimination, it means the terminal nodes of the key. The key can identify all or just a set of taxa, or any group of taxa defined by character state. So, considering a knowledge base describing species and a character “genus”, the user can then create a key discriminating the different genera and then keys to recognize species within each genus. In the same manner we can compute a key to identify the toxicity of mushrooms and not the species themselves.

Considering a knowledge base that covers a world distributed taxon, the user may need to consider only a subarea (hereafter called “sub-base”). The sub-base will then only include the taxa specified by the user. If the user can fill in a background (a specific region or country, a maximal bathymetric range etc.), a sub-base will be extracted, excluding taxa not compatible with the given conditions. The decision tree generated by MAKEY is then shorter than the key including all taxa, and so it minimizes the probability of error. Indeed, if two taxa are quite similar but are not present in the same altitude/country, using a key built on a sub-base reduces the risk of misidentification.

3.2 Background knowledge

Weights (or ponderation values), one for each descriptor, define a pre-order on the descriptor set (by default an equal weight is associated to all the descriptors or characters). MAKEY will respect this pre-order to select the character at each node. The characters can be ordered by the final user himself to force their choice in the key. So if flowers are absent the weight of all the flower's characters can be minimized or put to zero. At the contrary if some characters are easy to observe for the user he can associate to these characters a higher weight.

3.3 Topology of the identification graph

The topology section lets the user to choose some criteria to be used during the key construction (minimal number of branches at each node; to merge branches; to eliminate first some taxa etc.). Some statistics measurements help to compare the topology of the keys with different parameters and to choose the best decision tree.

3.4 Output format

The user can define the parameters to display the key: nested key (also called "yoked" or "indented") or parallel key (also called "bracketed" or "linked" key). Additional characters and states may be added if they are deduced at a step of the key.
The generated key is available in HTML format (including an option for a special layout for handheld devices) or in PDF for printing.

4 Architecture

Mykey is a server-side software implemented as a CGI script written in PYTHON; the system is easy to maintain and to upgrade, and it is compatible with any operating system.

According to the user selected parameters, (a) Mykey extracts a sub-base if necessary, (b) Mykey creates or modifies the file of character weights, (c) Mykey calls the software MAKEY which is then executed on the server with the selected parameters and (d) Mykey formats the MAKEY output and the result is sent to the client browser in the selected design. The key can also be saved on the server (in fact only the parameters will be saved), to restore it when needed, to modify it or to share it with other users.

5 Conclusion

Mykey is a running prototype. It is an efficient additional system to Xper^2, a midfield solution between single access key and free access key. Today a depository for Xper^2 knowledge bases is accessible at http://lis-upmc.snv.jussieu.fr/xper2/infosXper2Bases/en/index.php to any user. Then the data

can be accessible with Mykey. An option modifies the display for output on a personal pocket palm. Few similar options were encountered (URL: http://www.phylodiversity.net/palmkey/), and the one proposed by MyKey is perfectible.

Mykey is not a website to access to keys but an online service to produce keys [17]. In the European project EDIT the functions to create keys were implemented in the CDM library (see [20] in this book).

Mykey has to be modified to become a web service able to be connected easily to other softwares. In the future ViBRANT project (Virtual Biodiversity Research and Access Network for Taxonomy http://vbrant.eu) such identification system (free access and single access key construction) will be available as a web service and will allow a more open and flexible use.

Acknowledgement

The authors wish to thank Amandine Sahl for her contribution to this work during her master PhD, and all the users of this prototype.

References

[1] J. D. Agosti, "Biodiversity data are out of local taxonomists' reach". *Nature*, p. 392, 2006.

[2] J. R. Quinlan, "Induction of decision trees". *Machine learning,* vol. 1, pp. 81-106, 1986.

[3] D. W. Goodall, "Identification by computer". *Bioscience,* vol. 18(6), pp. 485-488, 1968.

[4] R. J. Pankhurst, "Identification methods and the quality of taxonomic descriptions". In: *Biological identification with computers*. Academic Press, London, 1975.

[5] P. M. Forget, J. Lebbe, H. Puig, R. Vignes and M. Hideux, "Microcomputer-aided identification / an application to trees from french Guiana". *Bot. J. Linn. Soc.*, vol. 93, pp. 205-223, 1986.

[6] M.J. Dallwitz, Overview of the DELTA System, 2009. http://delta-intkey.com/www/overview.htm, June 2010.

[7] M. J. Dallwitz, T. A. Paine and E.J. Zurcher, *User's Guide to Intkey: a Program for Interactive Identification and Information Retrieval,* vol. 1, 1995.

[8] J. Lebbe R. Vignes and J.P. Dedet, "Computer-aided identification of insect vectors". *Parasitology Today,* vol. 5 (9), pp. 301-304, 1989.

[9] A. R. Brach and H. Song, "eFloras: New directions for on-line floras exemplified by the Flora of China Project". *Taxon*, vol. 55 (1), pp. 188-192, 2006.

[10] R. J. Pankhurst, *Practical Taxonomic Computing*. Cambridge Univ. Press, Cambridge, 1991.

[11] J. Lebbe and R. Vignes, "State of the art in computer-aided identification in biology". *Oceanis*, vol. 24(4), pp. 305-317, 1998.

[12] K. Causse and J. Lebbe, "Modélisation des stratégies d'identification par la méthode MCC". *JAVA-95,* (Conference proceedings), 1995.

[13] J. Lebbe and R. Vignes, "Génération de graphes d'identification à partir de description de concepts". In: Y. Kodratoff and E. Diday (eds.), *Induction Symbolique et numérique à partir de données*, Cepadues, pp. 193-239, 1991.

[14] V. Ung, G. Dubus, R. Zaragüeta-Bagils and R. Vignes Lebbe, Xper²: introducing e-Taxonomy. *Bioinformatics,* vol. 26(5), pp.703-704, 2010.

[15] R. Vignes, *Caractérisation automatique de groupes biologiques*. Université Pierre et Marie Curie, 260 pp. (Thesis), 1991.

[16] J.C. Gower and R.W. Payne, "A comparison of different criteria for selecting binary tests in diagnostic keys". *Biometrika,* vol. 62, pp. 665-672, 1975.

[17] N. Conruyt, D. Sébastien, S. Cosadia, R. Vignes Lebbe and Touraïvane, "Moving from biodiversity information systems to biodiversity information services". In: L. Maurer, K. Tochtermann (eds.), *Information and Communication Technologies for Biodiversity Conservation and Agriculture*, Shaker, Aachen, (ISBN: 978-3-8322-8459-6), 2009.

[18] J. Nascimbene, S. Martellos and P. L. Nimis, "An integrated system for automatically producing

user-specific keys - A case study on Italian lichens". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 151-156, 2010.

[19] E. van Spronsen, S. Martellos, D. Seijts, P. Schalk, and P. L. Nimis, "Modifiable digital identification keys". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 127-131, 2010.

[20] W. G. Berendsohn, "Devising the EDIT Platform for Cybertaxonomy". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 1-6, 2010.

[21] G. Hagedorn, G. Rambold and S. Martellos, "Types of identification keys". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 59-64, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 113.
ISBN 978-88-8303-295-0. EUT, 2010.

Xper²: managing descriptive data from their collection to *e*-monographs

Visotheary Ung, Florian Causse, Régine Vignes Lebbe

Abstract — Computer Aided Identification (CAI) systems provide users with the resources to relate morpho-anatomic observations with names of taxa, and to subsequently access other knowledge about the organisms. Xper² version 2.1 is one of the most user-friendly software in its category. It provides a complete environment dedicated to taxonomic descriptions management. While assisting taxonomists with knowledge acquisition for identification keys, it also helps with the publication of descriptive data by providing a large panel of tools, including analysis facilities (comparison of taxa or descriptors used), full traceability of information and knowledge (by adding references or external links). Xper² provides excellent support for automatic on-line publication of descriptive data and free access keys, as well as for exporting of datasets for phylogenetic and systematic research. It focuses on interoperability between systems and can import and export into structured descriptive data format (TDWG-SDD), and export to HTML and Nexus formats. Written in Java, it is available on Windows™ Mac™ or Linux in French, English, and Spanish, and a new Chinese version can also be downloaded. With its intuitive interface, Xper² is aimed at professional taxonomists as well as naturalists who merely want to identify specimens using a ready-made application. Xper² is free of charge. It can be downloaded at: http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/softwares/xper2.

Index Terms — descriptive structured data, knowledge base, interactive identification key, integrated platform for identification.

V. Ung is CNRS engineer in UMR 7207 CNRS/MNHN/UPMC, MNHN Département Histoire de la Terre, CP48, 57 rue Cuvier, 75005 Paris, France E-mail: visotheary.riviere-ung@snv.jussieu.fr.
F. Causse is UPMC engineer in UMR 7207 CNRS/MNHN/UPMC, MNHN Département Histoire de la Terre, CP48, 57 rue Cuvier, 75005 Paris, France E-mail: florian.causse@upmc.fr.
R. Vignes Lebbe is Professor in UMR 7207 CNRS/MNHN/UPMC, MNHN Département Histoire de la Terre, CP48, 57 rue Cuvier, 75005 Paris, France.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 115-120.
ISBN 978-88-8303-295-0. EUT, 2010.

FRIDA 3.0 Multi-authored digital identification keys in the Web

Stefano Martellos

Abstract — FRIDA (FRiendly IDentificAtion) is a software for generating different types of digital identification keys starting from a database of characters. Its first release dates back to 2003. Since the first version, FRIDA was constantly developed to improve its core functions, and to add new features. At the beginning of 2010 it was decided to develop a new version, 3.0, written in PHP language for MySQL engines. The new version will have new and enhanced features, as well as the possibility to interact with other software developed in the framework of the European project *KeyToNature*. The development is currently in the beta-testing phase, and the final release is planned for the beginning of 2011.

Index Terms — biodiversity informatics, *Dryades*, *KeyToNature*.

1 Introduction

The first approaches to digital identification are recent, dating back to the '70s of the last century, and especially to the beginning of the "explosion" of the World Wide Web, less than twenty years ago. Today there is a great and continuously increasing number of different digital identification keys, produced by several research centers: fixed- or free-pathway keys, with different querying systems, matrix keys, simple textual keys, etc. [1], [2]. They can be accessible on the Web, stored on CD- or DVD-ROMs, and some of them can run on mobile devices as PDAs and Smartphones [3], [4], [5], [6], [7], [8], [9], [10].

The production of identification keys, both "classic" and digital, to large groups of organisms (e.g. a national flora), normally requires the combined effort of several authors. In most of the classic, paper-printed keys, each author (or small groups of authors) develops one or few keys to families, species or genera, which are then connected in a hierarchical way by a "general key". This approach has been succesfully applied to digital keys as well, e.g. in the Flora of China project

S. Martellos is with the Department of Life Sciences, University of Trieste, I-34127, Italy. E-mail: martelst@units.it.

[10], in which a different dataset for each family or genus was developed by one or a few authors. A different approach is that of many authors working together on a common dataset, e.g. in the project LIAS for lichenized and lichenicolous fungi [11]. Another interesting experiment to develop a community approach in building multi-authored digital keys is the BioWikiFarm [12]. In this case, the keys are digital texts stored on a MediaWiki platform, and a potentially very large community of authors can edit them, while the system registers and keeps track of the changes.

FRIDA (FRiendly IdentificAtion) [13] was developed to allow several authors to work together, but rather independently from each other, while building a common database of morpho-anatomical data from which it is possible to produce a virtually unlimited number of different multi-authored digital keys. This software has been already used to generate hundreds of keys to plants, animals and fungi in the framework of the European project *KeyToNature*. FRIDA was available as a package running on Oracle databases only up to its last version (2.0). The new version, FRIDA 3.0, currently in beta stage of development, is written in PHP for MySQL databases, and has several new and improved features. Furthermore, it will be possible to use it both in stand-alone and on-line mode.

2 FRIDA

FRIDA (FRiendly IDentificAtion) has been developed since 2003 at the Department of Life Sciences of the University of Trieste (Italy), in the framework of project *Dryades* [7]. Up to its version 2.0 it was written in PL/SQL language, and developed on an Oracle Database engine.

The most interesting features of FRIDA [13] are:

1. *It does not require the learning of any code or programming language.* Input and management of data are in natural language, through simple Web interfaces written in HTML 4.0.
2. *Keys are immediately available on-line since their generation.* They are accessible from the Web by using any common web browser, through a single-access and a multi-entry query interface [8].
3. *Keys are independent from the original data.* When a key is produced, it does exist as a discrete entity, separated from the original database. In this way it is possible to modify the keys whithout affecting the original database, or vice-versa.
4. *The database of characters has a double-level architecture.* Characters are stored in two levels of information: a) a first level which is common to all taxa in the database, b) a second level, which is restricted to taxa belonging to a given "group" (see later). Organisms can be divided into more homogenous groups (e.g. genera and families, but also fully artifical groupings) by using several characters of the first level. These groups do exist as independent entities in the second level, and can be managed as independent databases by different authors (e.g. specialists of a given genus or family), which can thus work with a large degree of independence.
5. *The weight of characters in the generation of dichotomous keys is decided*

by the authors case by case. While it is possible to use an algorithm [14] to produce the "better" key (e.g. the key with the shortest branches), only an experienced taxonomist knows which is the weight of a given character in a particular group of organisms.

6. *Keys are portable in the field, both online and in stand-alone versions.* The latter, while less performing, are the better solution when internet connections are not available or poorly effective. Stand-alone keys can be also stored on CD- and DVD-ROMs.

3 New features in version 3.0

The development of FRIDA 3.0 started at the beginning of 2010. The new version is written in PHP language and runs on MySQL engines. It has the same core features of the previous version, plus some new and enhanced features:

1. *It can be also installed and used locally*. While the previous version required an Oracle Database and an Application Server to be installed and could be used on-line only, this version can run locally on any computer provided with MySql, PHP and Apache. It can be obviously installed on a server computer as well.
2. *Enhanced multilingual support*. While in the previous version it was possible to input the data into two languages only, in the new version the number of languages can be decided by the authors, and is virtually unlimited. The languages are selected in the settings interface, and new languages can be added at any time.
3. *New templates.* The development of the previous version focused almost exclusively on software performance, the interfaces for data management were rather poor and sometimes tricky for the users. In the new version, thanks to a new template, the functions are accessible in a far more rational way. For example, the management of characters can be done by using a few interfaces, which group all the character and states functions (Fig. 1), while the old version had a different interface for each function.
4. *Improved management of character images.* In the previous version the images of characters could be stored and reused, but the process was tricky, and sometimes forced the authors to create "dummy" characters with several fake states. In the new version character images can be stored in a image archive, and searched by file name or keyword.
5. *Easier management of records.* In FRIDA each taxon can be described by several records. Each record, while having the same name, and referring to the same taxon, differs for at least a character state. In the new version all the records referring to a taxon can be viewed together, and managed in a more efficient way (Fig. 2).
6. *Commented functions.* All the interfaces are provided with short textual information, to give the authors reminders on the use of the different functions. A full reference manual to FRIDA 3.0 is in preparation.

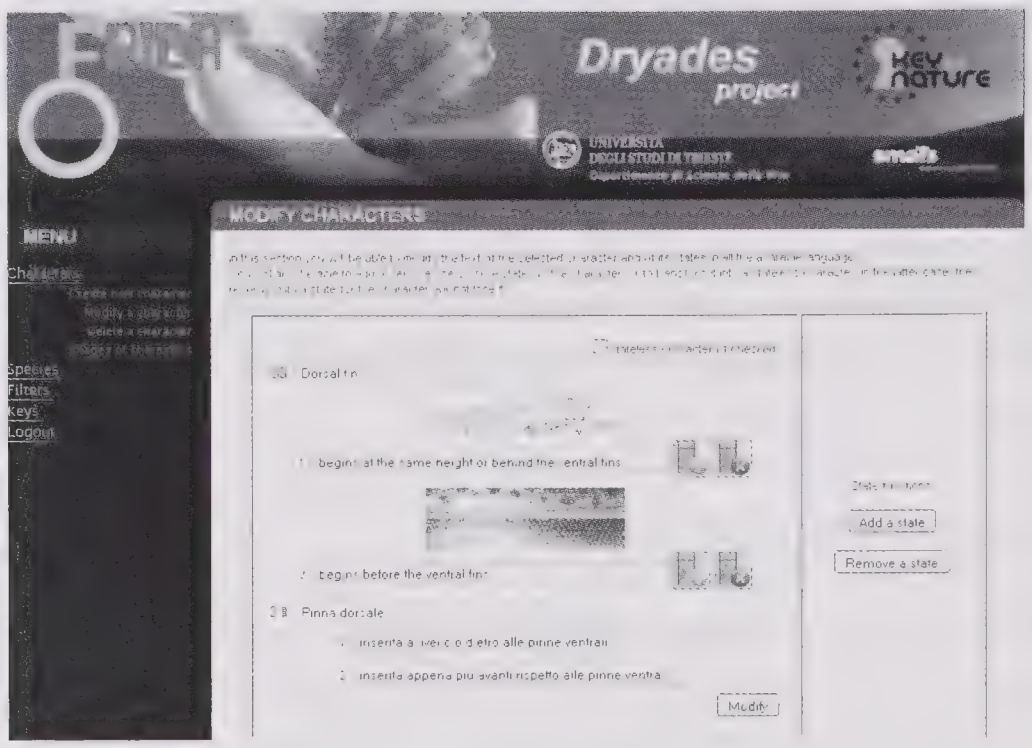


Fig. 1 – The new interface for the management of characters and values. Several functions are grouped together.

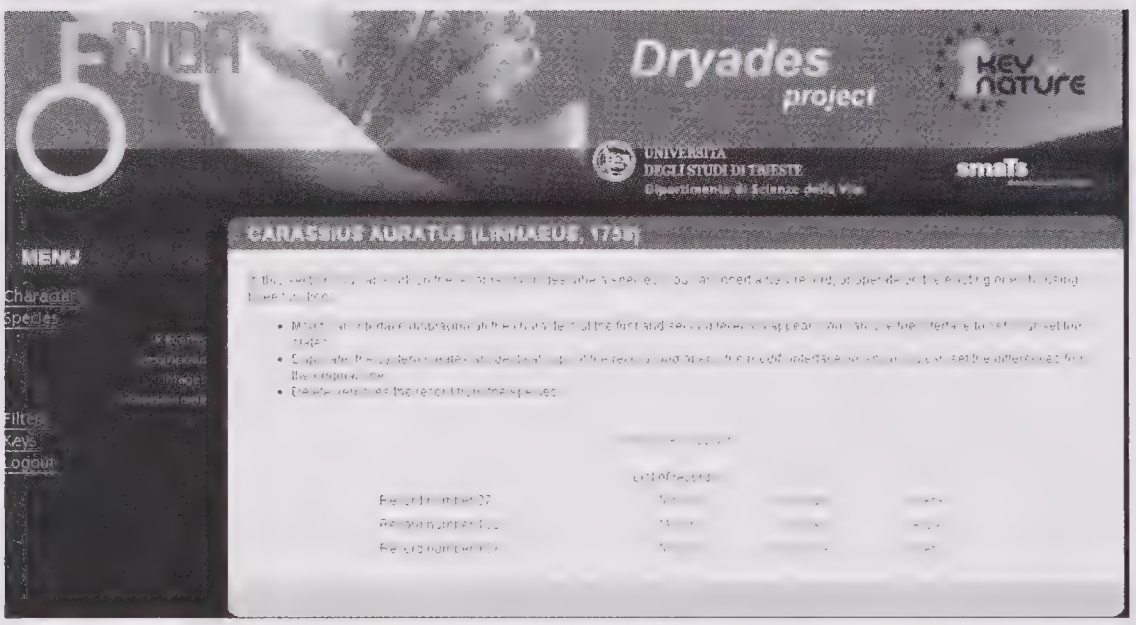


Fig. 2 – Each taxon can be described by several records, differing for at least a character state. The records are managed in a simple interface, which permits to edit, duplicate and delete them, as well as to add new records.

4 Conclusion

FRIDA 3.0 will be accessible to several research centers, including those without an Oracle system. It will permit to export the keys both in the Open Key Editor [15], [16], [17], in the Open Key Player [18] and in the BioWikiFarm [12] formats, which were developed in the framework of *KeyToNature* [19], to contribute to the development of integrated, open networks for digital identification.

The estimated roadmap for the future development of FRIDA 3.0 is:

- November, 2010 – FRIDA 3.0 Beta 2,
- December, 2010 – FRIDA 3.0 Beta 3,
- January, 2011 – FRIDA 3.0 Release Candidate (RC) 1,
- February, 2011 – FRIDA 3.0 RC 2,
- March, 2011 – FRIDA 3.0 - official release.

While the beta testing phase is closed, the RC versions will be available upon request to the author.

Acknowledgement

This paper was produced in the framework of the project *KeyToNature*, funded under the *e*Content*plus* programme, a multi-annual Community programme to make digital content in Europe more accessible, usable and exploitable. — Contract no. ECP-2006-EDU-410019.

References

[1] M. J. Dallwitz, T. A. Paine, and E. J. Zurcher, *Principles of interactive keys.* (http://delta-intkey.com), 2000 (onwards).

[2] M. J. Dallwitz, T. A. Paine and E. J. Zurcher, *Interactive identification using the Internet.* (http://delta-intkey.com), 2002 (onwards).

[3] G. Agarwal, H. Ling, D. Jacobs, S. Shirdhonkar, W. J. Kress, R. Russell, P. Belhumeur, A. Dixit, S. Feiner, D. Mahajan, K. Sunkavalli, R. Ramamoorthi and S. White, "First steps towards an electronic field guide for plants." *Taxon,* vol. 53 (3), pp. 597-610, 2006.

[4] A. R. Brach and H. Song, "ActKey: a Web-based interactive identification key program". *Taxon,* vol. 54 (4), pp. 1041-1046, 2005.

[5] D. F. Farr, "On line keys: more than just paper in the web". *Taxon,* vol. 53 (3), pp. 589-596, 2006.

[6] K. Chang-Sheng and H. Song, "Interactive key to Taiwan grasses using characters of leaf anatomy – the ActKey approach". *Taiwania,* vol. 50, pp. 261-71, 2005.

[7] S. Martellos and P. L. Nimis, "*KeyToNature*: Teaching and Learning Biodiversity. Dryades, the Italian Experience". In: M. Munõz, I. Jelínek, and F. Ferreira (eds.), *Proceedings of the IASK International Conference Teaching and Learning 2008*, pp. 863-868, 2008.

[8] R. D. Stevenson, W. A. Haber and R.A. Morris, "Electronic field guides and user community in the ecoinformatics revolution". *Conservation Ecology,* vol. 7(1), pp. 3, 2003

[9] M. J. Dallwitz, "*A comparison of interactive identification programs*". (http.//delta-intkey.com), 2000 (onwards).

[10] A. R. Brach and H. Song, "eFlora: New directions for online floras exemplified by the Flora of China Project". *Taxon,* vol. 55(1), pp. 188-92, 2006.

[11] G. Rambold, "LIAS – The concept of an identification system for lichenized and lichenicolous fungi". In: Anonymous (ed.), *The Third Symposium IAL 3. Progress and problems in Lichenology in the Nineties. Abstracts. - 9. Salzburg*, 1996.

[12] G. Hagedorn, G. Weber, A. Plank, M. Giurgiu, A. Homodi, C. Veja, G. Schmidt, P. Mihnev, M. Roujinov, D. Triebel, R. A. Morris, B. Zelazny, E. van Spronsen, P. Schalk, C. Kittl, R. Brandner, S. Martellos, and P. L. Nimis, "An online authoring and publishing platform for field guides and identification tools". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity. Progress and Problems,* pp. 13-18, 2010.

[13] S. Martellos, "Multi-authored interactive identification keys: the FRIDA (FRiendly IDentificAtion) package". *Taxon*, vol. 59 (3), pp. 922-929, 2010.

[14] M.J. Dallwitz, T.A. Paine and E. J. Zurcher, "*User's guide to the DELTA System: a general system for processing taxonomic descriptions*". 4th edition. (http://delta-intkey.com), 1993 (onwards).

[15] S. Martellos, E. van Spronsen, D. Seijts, N. Torrescasana Aloy, P. Schalk and P. L. Nimis, "Digital identification keys to organisms and user-generated content. The *KeyToNature* approach". In: M. Muñoz and F. Ferreira (eds.), *Proceedings of the IASK International Conference Teaching and Learning 2009*, pp. 96-102, 2009.

[16] S. Martellos, E. van Spronsen, D. Seijts, N. Torrescasana Aloy, P. Schalk and P. L. Nimis, "User-generated content in the digital identification of organisms: the *KeyToNature* approach". *Int. J. Information and Operations Management Education*, vol. 3, 3, pp. 272-83, 2010.

[17] E. van Spronsen, S. Martellos, D. Seijts, P. Schalk and P. L. Nimis, "Modifiable digital identification keys". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems,* pp. 127-131, 2010.

[18] M. Giurgiu, A. Homodi, E. van Spronsen, S. Martellos and P. L. Nimis, "The Open Key Player:

A new approach for online interaction and user-tracking in identification keys". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems,* pp. 133-136, 2010.

[19] G. Hagedorn, P. L. Nimis and P. Schalk, "*KeyToNature*: Software, data formats, and communities". *Biodiversity Informatics Symposium 2008. The Book of Abstracts*, Swedish Museum of Natural History, Stockholm, Sweden, 27, 2008.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 121-125.
ISBN 978-88-8303-295-0. EUT, 2010.

Flora Bellissima, an expert software to discover botany and identify plants

Thierry Pernot, Daniel Mathieu

Abstract — To promote and facilitate access to botany for protecting biodiversity indirectly is the objective of Flora Bellissima. This sofware package is based on an integrated and complete database with numerous fonctionalities, as well as on an expert system of recognition of plants called "*Ophélie*". Aimed at beginners, amateurs and experts, Flora Bellissima attempts to show that it is possible to reconcile scientific rigor and popularization, in order to bring together everyone interested in botany.

Index Terms — botanical data management, expert system, plant recognition help, education through play.

1 Introduction

How to raise awareness in the public of what biodiversity is and of its protection? A solution is to help them to discover the nature around them, so they can know and protect it better. To acquire this knowledge, people need tools which are fun, easy, and based on scientific knowledge. The tools must be also adaptable to different audiences, from novices to experts, and they must take into account the progression of each individual's learning.

It is in this perspective that Flora Bellissima [1], an integrated management sofware dedicated to botany, has been developed. With a organisation similar to a " mini ERP" (Enterprise Resource Planning), this software is based on a central database composed of a scientific index of plants (BDNFF [2]). Four modules are organised around this database. They allow : (i) an educational discovery of the flora, (ii) assistance in plant determination thanks to the expert system « *Ophélie* », (iii) a management system for botanical data collected during field trips, and, (iv) a game module.

T. Pernot, the autor of the software, is with Yourproject Informatique 27 rue Saint-Georges 39360 Larrivoire France Yourproject@orange.fr - www.yourprojectinfo.fr. He is a biologist by training (Master of biology of organisms and populations at the University of Besançon) and computer scientist (project manager in corporate and SSI).
Daniel Mathieu is the president of Tela Botanica, 163 rue A. Broussonet, 34090 Montpellier, France, the NGO which publishes and distributes the sofware. dmathieu@tela-botanica.org www.tela-botanica.org.

This software has been developed for anybody interested in the flora, whatever their knowledge level. Novices who wish to make their first steps in botany can learn in an entertaining way thanks to the game module, the image glossary and the photographs. Amateurs who wish to improve their knowledge can store the result of their observations (texts and photographs) and build up their own database. Experts can consult easily and quickly the botanical nomenclature with all its synonyms and its taxonomic levels : species, sub-species, varieties and forms. The naturalist organisations can quickly save and exchange their field botany notes. Everybody can use the expert system *Ophélie* to identify a plant and give it a name before recording it.

2 Tool conception

2.1 General specifications

The detailed analysis of the different objectives led us to make the choices found in the main concepts of Flora Bellissima:

- the software is made like an ERP, i.e. composed of several applications sharing the same and unique database ;
- It is based on a scientific nomenclatural referential index ;
- It is open to allow the addition of text and photographs ;
- It is multi-level to be accessible to anyone : novices, amateurs and experts ;
- It contains data on the French flora for 1 400 species ;
- It proposes a help tool to determine plants, adapted to all audiences with the expert system *Ophélie* ;
- It has several photographs for each plant (9 800 photograhs in total with general appearance, inflorescence, flower, leaves etc.) to validate the determination ;
- It is available on PC and is distributed on DVD.

2.2 The flora discovery module

This module helps the discovery of the flora of France through numerous photographs and allows users to build up their own database based on their own observations. It also includes more than 30 functionalities which are not detailed here.

In summary, this module allows the consultation of a flora depending on different points of view : plant type (tree, shrub, fern, etc.), plant use (medicinal, cultivated, toxic), botanical family, genus, Latin or vernacular name. It also allows the consultation and capture of information on different themes : description, medicinal properties, protection, and geographical distribution [3]. Note that it is possible to add and link your own photograhs to each taxon. The three access levels, the picture glossary and the ergonomy give a strong educational value to this module.

The important idea emerging from this module is the gathering and centralisation of information in order to facilitate the access to it.

2.3 The « plant station » note management module

This module allows “plant stations” to be defined with their geographical and ecological characters, then to connect them to “botanical field trip notes”. This tool has been designed for capturing efficiently the complete Latin names of plants. Besides its functionalities of classification of stations and botanical field trip notes, this module proposes the following options :

- plant search among all field trip notes ;
- analysis of the evolution of the plant population of a station through time ;
- copying field trip notes ;
- import/ export of stations and field trip notes ;
- printing field trip notes and exporting them as a Word, Excel, PDF etc file.

2.4 The determination aid module

This module is made up of two main functions : a system of comparison between plants and an expert system of determination.

The comparison system allows the comparison of between 2 to 5 plants displaying all the characters which can differentiate them. This quickly gives the important characters to differentiate plants from each other. The system, called *Ophélie*, is an original tool which helps in the determination of plants using the Flora Bellissima database. One thousand four hundred species are now described, with about 30 characters for each plant, i.e. 45 000 characters described for the whole database.

To make sure that the system is efficient, it is important to take into account two main constraints: i) not all users are experts, and they can make mistakes, ii) plant morphology differs from one plant sample to another. Also, to work properly, the system must be able to accept mistakes and variations in the descriptions.

The system can work only on a binary way with the characters: true or false. It must work on all the characters described simultaneously, but not one after the other, as it is the case with a dichotomous key.

Each entered value allows a “weighting factor” to be allocated to all possible values of a character. To do this, we need to know “the distances” separating the different values of the same character. This is done by the similarity matrices which represent over 11 000 datapoints in Ophélie. In order to be able to enter such a quantity of datapoints, specific computing tools have been developed. These tools enable data capturing to be speeded up while checking it at the same time. Ophélie can manage different types of characters such as matrix, interval, quantitative values, etc.

The “weighting factors” are allocated to each of the chosen values thanks to algorithms which depend on the character type: for instance, if we input the value “pink” to the colour of a flower, this value will have the highest weighting factor, the value “red” would have a lower weighting factor, and the value “yellow” an even lower value...

In order to limit the number of questions to which the user must answer, Ophélie has an algorithm which optimizes the path to the final answer (plant name) by

taking into account the previous answers. The system can determine which criterion - until then unused in the process - can provide the best information. Anything complex is handled by the computer, not by the user!

Dealing with the large number of descriptive characters to fill in - approximately 700 for 1 400 species, is another issue which had to be worked out. It was obviously not conceivable to enter all this information manually for each species!

In Ophélie, the solution was to introduce hierarchical levels of description which allow the factorization of descriptions. Three hierarchical levels are used: (i) "general" to differentiate families, (ii) "family" to differentiate genera and (iii) "genus" to differentiate species. This arrangement has led to a great reduction in the number of descriptive characters to type in, from 700 to approximately 30.

Note that the system works in a global way and that consequently, the "general" level can be enough for the determination of some species! However, the management of level changes is a delicate step. To do so, two solutions are used:

- delay the moment we activate the level change, but not excessively in order to maintain a reasonable number of questions;
- checking descriptive characters for level change.

Once the determination is finished, the system *Ophélie* only proposes a plant name if the difference between this plant and other plants is big enough, and if the degree of similarity with the description of the given plant has reached a certain level. This allows the reduction of the risk of getting an incorrect determination. The quick display of all photographs concerning all plants corresponding closely to the description is also a good control to validate the result.

The *Ophélie* system is therefore based on the principle of separation of species, *i.e.* individuals of the same species are generally more similar than individuals from related species.

With this structure, the expert system *Ophélie* could work with a number of species superior to 1 400, and must be able to take into account the 6 000 species of the French flora. The performance of this system is solely dependent on the quality of the information of the database and the sharpness of the parameter setting.

In conclusion, *Ophélie* solves the problem of absence of an answer which occurs when using a determination key and which stops the determination. Also, with *Ophélie*, a bad choice for a few answers is not a major problem for the final determination, because it is the mean of all the answers which is important.

2.5 The game module

This module proposes two games with three levels (Novice/ amateur/ expert). Their aim is to make users learn how to recognize plants in an entertaining way through the detailed observation of photographs..

3 Conclusions

Conceived in order to take action on the problems of biodiversity in a sustainable and indirect way, Flora Bellissima is an attempt to bring together

different categories of botanists : novices, amateurs and experts in proposing an educational and entertaining software based on strong scientific knowlegde.

Flora Bellissima represents 4 000 hours of work including both the software design and the setting up of the knowledge base. Its extension to the whole French flora will require the setting up of a collaborative working group in order to capture characters concerning all taxa. The organization of such a group will be managed by the Tela Botanica association, which groups together all French-speaking botanists, and which has the qualifications and abilities required to achieve such a complex task. Tela Botanica also plans, in the long term, the online consultation of Flora Bellissima and could propose its application to other floras in the world within the framework of the research project Pl@ntNet [4] as a counterpart of existing systems.

Acknowledgement

We wish to thank the Tela Botanica team for their help in the realisation of this project as well as for its diffusion. We also want to thank Paul Fabre for translating this presentation in English.

References

[1] Flora Bellissima is a registered trademark of the Yourproject Informatique company.

[2] BDNFF: “Base de Données Nomenclaturale de la Flore de France” conducted by Michel Kerguélen † and Benoit Bock, in the Tela Botanica network.

[3] Maps of geographic distribution of plants are provided by Tela Botanica.

[4] Pl@ntNet is an Interactive plant identification and collaborative information system supported bay Agropolis Internatioonal Montpellier, France, built around three core teams that possess complementary skills : AMAP, IMEDIA and Tela Botanica.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 127-131.
ISBN 978-88-8303-295-0. EUT, 2010.

Modifiable digital identification keys

Edwin van Spronsen, Stefano Martellos, Dennis Seijts,
Peter Schalk, Pier Luigi Nimis

Abstract — The Open Key Editor (OKE) is a tool for editing and enriching existing identification keys and to produce localized 'minikeys' that apply to local flora and fauna, such as in parks, nature reserves and school gardens, or keys that apply to a particular season. The minikeys are easier to use than their originals, simply because of the fact that they deal with less species, their language can be adapted to a particular audience (e.g. pupils), and because they always point to species that are known to be present. OKE also allows the inclusion of user-generated content in any minikey (new text, images, hyperlinks etc.). The output of minikeys can be automatically tailored for display on computers, smartphones or PDA's.

Index Terms — biodiversity Informatics, field guides, identification tools, identification software, flora, fauna.

1 Introduction

Identification keys are often written by experts and aim at an 'academic' audience. Once they are published, they are more or less carved in stone and leave little room for adaptation to a specific audience, a particular region or season. In the case of plants, such a key can - for instance - encompass 1900 species for the Netherlands or more than 6000 species for e.g. Spain or Italy. Long keys are complicated and have redundant information when used in a region with fewer species, such as a park or nature reserve, or even a school garden. An increasing number of identifications tools is being published on the internet [1], [2], [3] offering an opportunity for tailoring them to particular audiences and situations. The Open Key Editor [4] is a software package developed within the *KeyToNature* European project, that allows users to 'crop' a master key and customize it for a given set of species. The 'cropped' key can then be edited for language and illustrations (e.g. to suit a particular user level, or platform such as the mobile phone).

E. van Spronsen, D. Seijts and P. Schalk are with ETI BioInformatics, Amsterdam. E-mail: edwin@eti.uva.nl, dennis@eti.uva.nl, peter@eti.uva.nl.
S. Martellos and P. L. Nimis are with the Dept. of Life Sciences, Univ. of Trieste, via Giorgieri 10, I 34127 Trieste, Italy. E-mail: martelst@units.it, nimis@units.it.

2 The Open Key Editor

The *KeyToNature* Open Key Editor is an easy-to-use Open-Source tool for editing and enriching a key with user-generated content. It was developed since June 2009, is written in the PHP 5.2 language, and runs on a MySQL 5.0 database. The code is Open Source and available under the Creative Commons Attribution Non-Commercial (CC-BY-NC) license. The program can import *dichotomous* and *polytomous* keys with a compatible structure. It is downloadable since December 2009 from the Web Portal of project *KeyToNature* (http://www.keytonature.eu), together with sample keys. The current version is 1.1.

With the Open Key Editor the user can browse existing master keys and edit them. The first step in making a customized 'mini-key' is to create a filter: this is a list containing a subset of the species of the main key. Such a list can be made by selecting species from the original key, or by importing a text file with species names from an external source. The filters can be stored for later editing, so many mini-keys can be tailored from the same basic dataset. In the Open Key Editor new couplets can be added to the key for identifying species that are absent in the original key.

The unprocessed mini-key will contain three kinds of questions from the original key:

1. valid ones that still separate (groups of) species.
2. questions that used to be like type 1, but now have only a single remaining branch.
3. questions that no longer lead to any species at all.

Questions of type 2 and 3 will have to be removed. Once a filter is defined, the programme starts with the species that were removed and traces them back until in encounters a question that is still relevant. All questions downstream are 'dead wood' (type 3) and will be removed. The application repeats this process with the remaining species in order to find questions of type 2. When it encounters a question that no longer separates at least two species, the question is removed from the decision tree, but its parent and child questions are connected in a new branching pattern. Because there is a chance that this new branching pattern will also contain questions of type 2, the whole process is reiterated until no more changes have to be made. The result is a key in which only questions of type 1 remain.

Special problems are reticulated keys. These are keys in which a question branches to another part of the key that is not 'downstream' of the present node. This problem is solved by controlling the creation of loops and unravelling them during the processing of the key.

2.1 Key viewer

The *Editor* is provided with a simple single-access query interface (Fig. 1), which displays, for each step of the identification process, one question and all its possible answers, enriched by images when available. Users can retrieve a list of the remaining species at each step of the identification process.

Furthermore, they can produce a printable, illustrated, identification key to the remaining species.

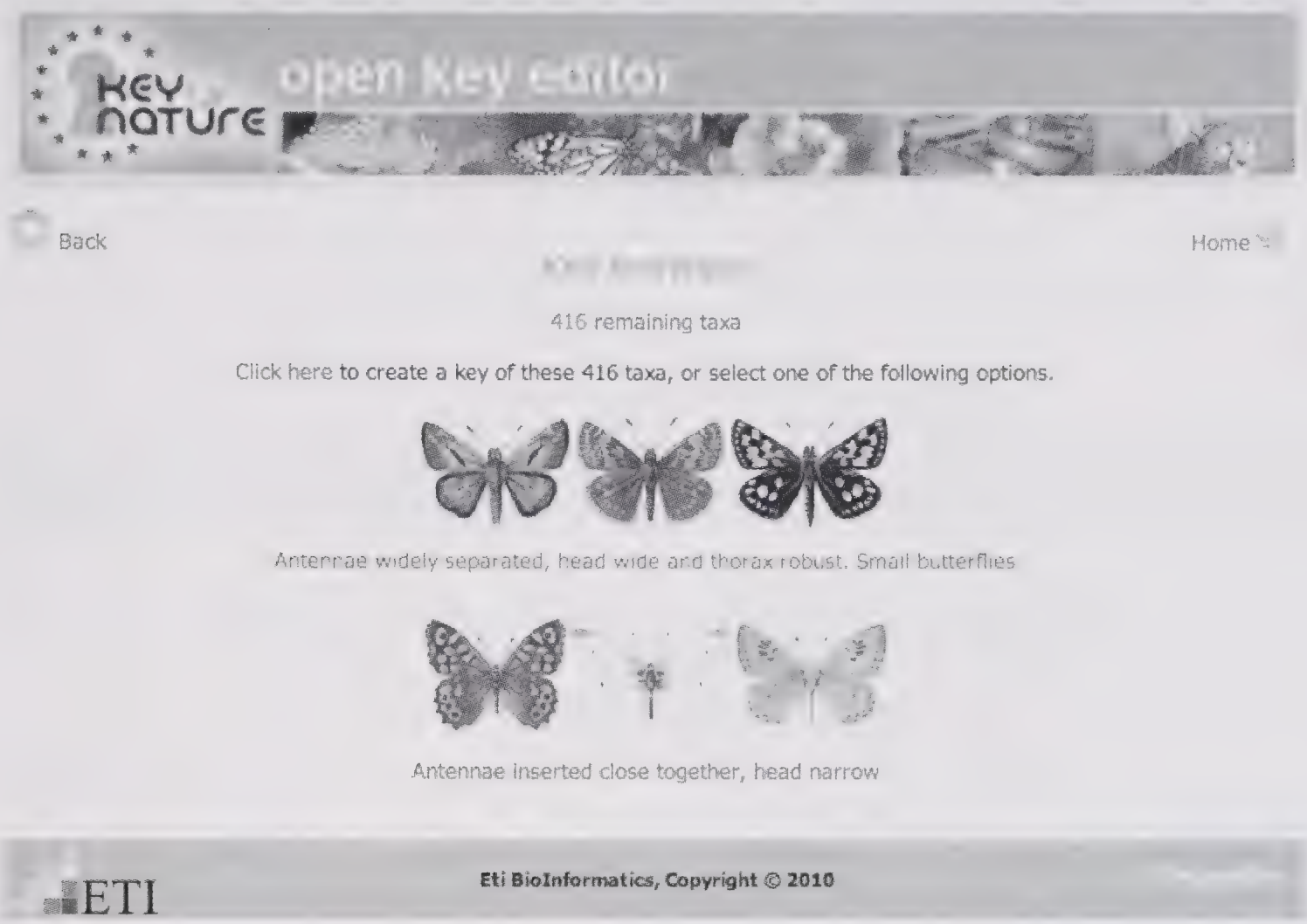

key nature
open key editor
Back
Home
416 remaining taxa
Click here to create a key of these 416 taxa, or select one of the following options.
Antennae widely separated, head wide and thorax robust. Small butterflies
Antennae inserted close together, head narrow
ETI
Eti BioInformatics, Copyright © 2010

Fig. 1 – A key as displayed by the Open Key Editor. At any time a minikey, based on the remaining species, can be generated.

2.2 Editing interface

The main functions of the Editor are devoted to the integration of user-generated content in the keys. Users can modify, or add user-generated content, to:

- the text of the key
- the pictures
- the names of the species, their images and descriptions, and links to external pages of interest

The editing process involves one line of a key at each time. The available options are displayed in a simple graphic interface (Fig. 2).

Apart from editing keys at the level of lists of species, a second level of editing is available: all texts of a key and its species descriptions can be changed. Jargon can be removed for pupils or added for specialists. Photographs or drawings can be uploaded by the user. They can be added to the original illustrations or even replace them. Changes in texts or illustrations of a minikey will not interfere with the original masterkey.

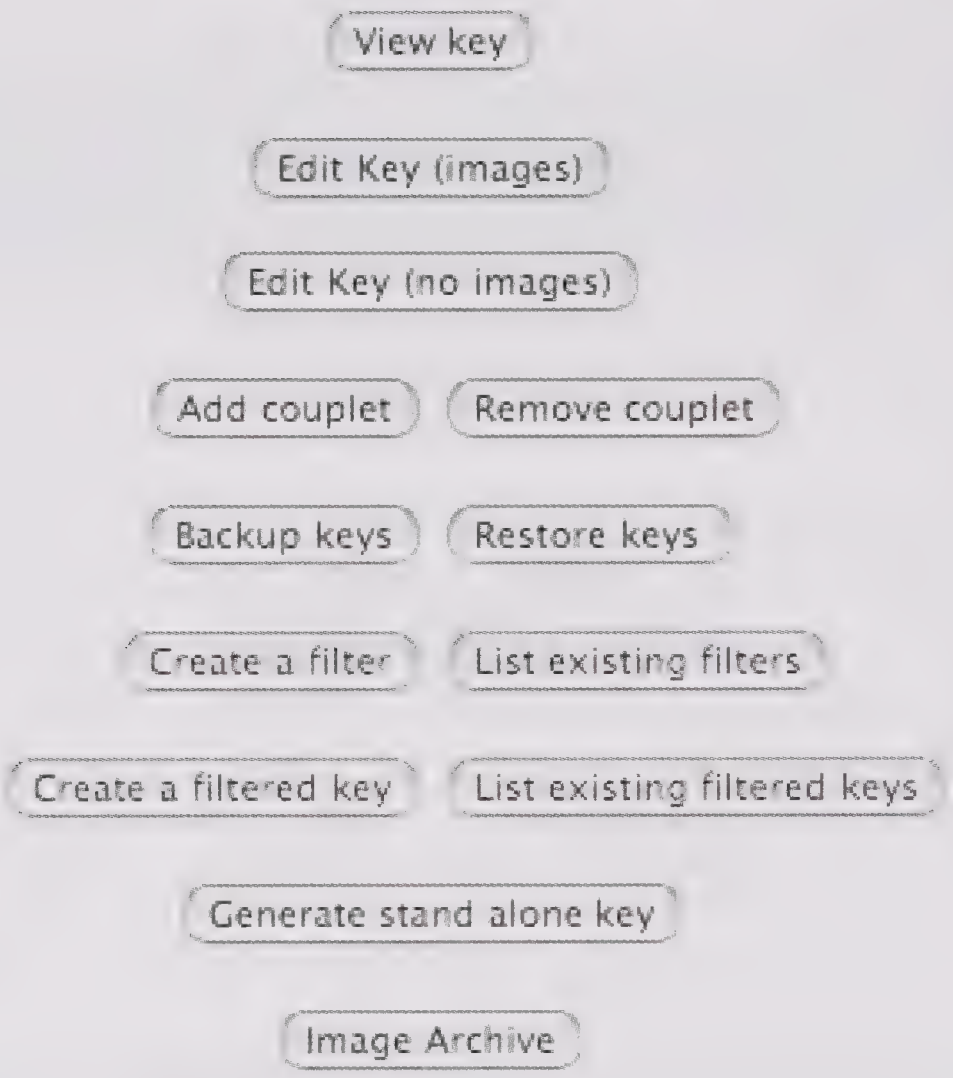


Fig. 2 – The administrative menu of the Open Key Editor.

2.3 Generation of 'filtered' keys

The Editor can generate ex-novo a virtually unlimited number of "filtered" keys from a single "master" key. After filtering, the "cropped" key becomes a separate entity, which can be edited independently from its "master" counterpart. A filtered key can be made available on the web, or given to a user for further editing. Any changes in texts or illustrations of a filtered key will not interfere with the original master key.

2.4 Export of *stand-alone* keys

Identification keys can be used at home or in a laboratory, but they can also be used in the field, while exploring the biodiversity of an area. The Editor can export the master key, or any of the filtered keys, in the form of stand-alone packages (Fig. 3). The stand-alone versions can be published on CD- or DVD-ROMs, or used on mobile devices, such as PDA's and smartphones. These devices, when equipped with a camera, can also be used to enrich the key with original pictures.

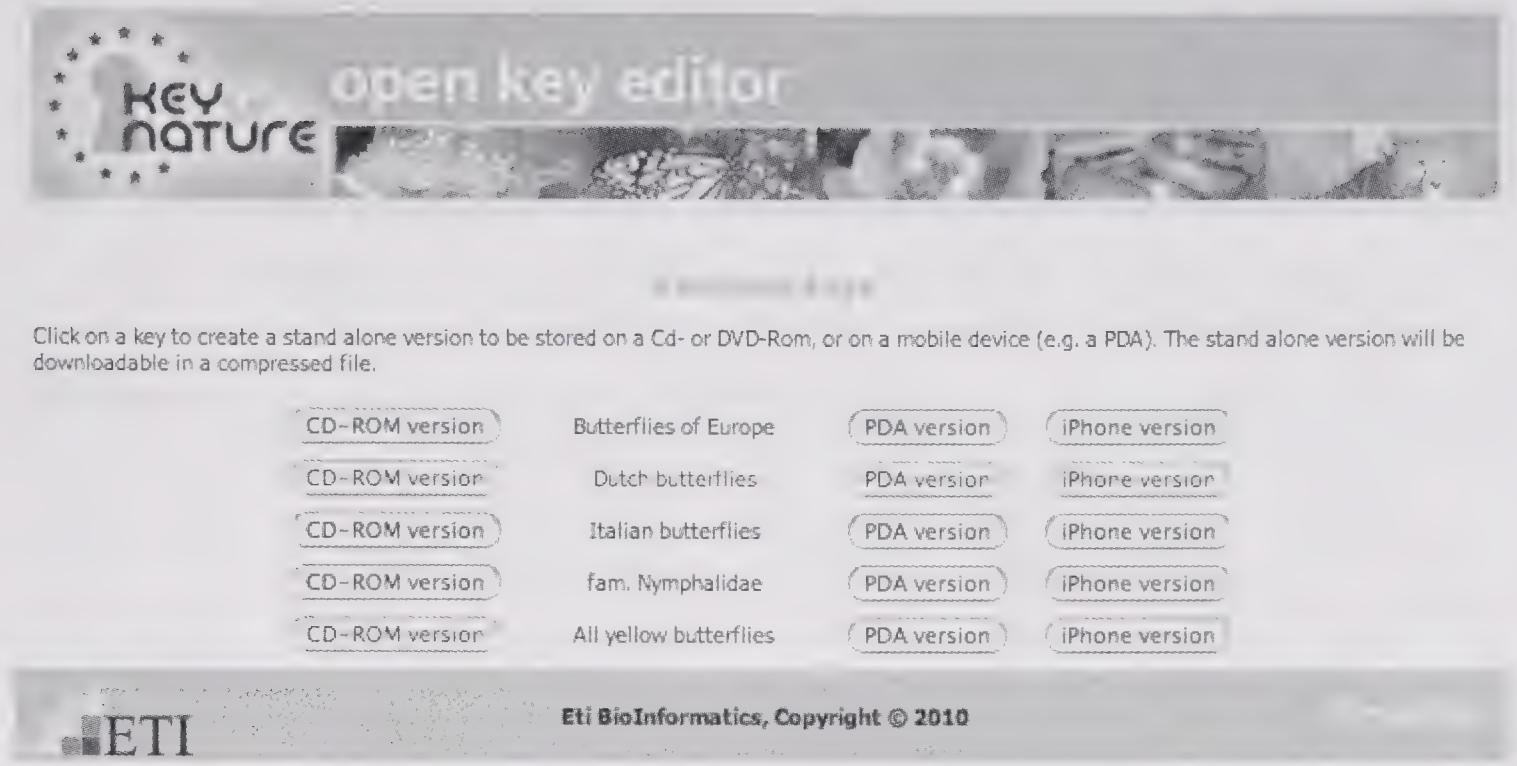


Fig. 3 – A single masterkey can produce different stand-alone minikeys, based on scientific or arbitrary criteria.

3 Conclusion

With the Open Key Editor, existing identification tools can be modified and their use can be made much easier by removing species that are absent from a particular region or season. Text of keys and species descriptions can be edited or translated so as to adapt them to user groups like pupils. Modified keys can be turned into stand-alone applications for computers, websites, smartphones or PDA's.

Acknowledgement

The authors wish to thank all the persons involved in *KeyToNature* throughout Europe. Their efforts and input gave us new ideas and energy to develop them. This paper was produced in the framework of the the project *KeyToNature* (www.keytonature.eu, ECP-2006-EDU-410019), funded in the e*Content*plus Programme.

References

[1] H. H. Visser and H. Veldhuijzen van Zanten, "European Limnofauna" http://ip30.eti.uva.nl/bis/limno.php?menuentry=sleutel, 2010.

[2] ETI Bioinformatics, "World Biodiversity Database" http://ip30.eti.uva.nl/bis/projects.php, 2010.

[3] S. Martellos and P. L. Nimis, "*KeyToNature*: Teaching and Learning Biodiversity: Dryades, the Italian Experience" In: M. Muñoz, I. Jelinek and F. Ferreira (eds.), *Proceedings of the International Association for the Scientific Knowledge (IASK) International Conference "Teaching and Learning",* Aveiro, Portugal pp. 863-868. (http://www.dryades.eu), 2008.

[4] S. Martellos, E. van Spronsen, D. Seijts, N. Torrescasana Aloy, P. Schalk and P. L. Nimis: "User-generated content in the digital identification of organisms: the *KeyToNature* approach". *International Journal of Information and Operations Management Education* (IJIOME), vol. 3, 3, pp. 272 -283, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 133-136.
ISBN 978-88-8303-295-0. EUT, 2010.

The Open Key Player: A new approach for online interaction and user-tracking in identification keys

Mircea Giurgiu, Andrei Homodi, Edwin van Spronsen,
Stefano Martellos, Pier Luigi Nimis

Abstract — This paper describes a new approach for creating and using identification keys based on two main components: the Open Key Editor and the Open Key Player, both of which were created within the European project *KeyToNature*. The Open Key Editor can be used to produce custom identification keys starting from a master key and to add original user-generated content, while the most important feature of the Open Key Player is the possibility to track relevant user activities into an eLearning environment, in order to collect data to improve the design and usability of identification keys.

Index Terms — identification key, eLearning, user-tracking, Rich Internet Application.

1 Introduction

Identification keys are used to identify biological entities such as plants and animals. Since the beginning of the digital era, the keys have undergone a great improvement from the early, paper-printed versions [1]. Modern digital keys are easier to access and to use, and can be used into schools as new, efficient and interactive instruments for teaching biodiversity. However, building an original identification key is still a task which can be carried on by an expert only.

M. Giurgiu and A. Homodi are with the Telecommunications Department, Technical University of Cluj-Napoca, Cluj 400027, Romania. E-mail: Mircea.Giurgiu@com.utcluj.ro.
E. van Spronsen is with ETI Bioinformatics, Amsterdam, E-mail: edwin@eti.uva.nl.
S. Martellos and P. L. Nimis are with the Department of Life Sciences, University of Trieste, I-34127, Italy, E-mail: nimis@units.it, martelst@units.it.

2 The Open Key Standard

The Open Key Standard from the *KeyToNature* project has been developed to make the creation and the use of an identification key easier, as well as to improve its accessibility (Fig. 1). The standard contains two major components: The Open Key Editor (OKE) [2], [3] and the Open Key Player (OKP).

The Open Key Editor provides the necessary interfaces to manage an identification key created from scratch, or imported using the Structured Descriptive Data (SDD) standard, as well as other file formats. From an original key, called "master key", it is possible to create a virtually unlimited number of derived keys to different lists of taxa, containing new and original user-generated content, and devoted to different target users.

The identification keys can then be used in the Open Key Player. This is a Flash application developed by using Adobe Flex, which operates on the database of the OKE, and displays in a modern, interactive way the keys to the user [4], [5].

3 Open Key Player Design

The Open Key Player has been designed following several requirements: 1) cross platform and browser compatibility, with the possibility to interact with MySQL databases, and to display data in XML (eXtendable Markup Language) format [6], [7]; 2) user interaction and information specific for dichotomous keys; 3) tracking the features selected along the identification path, and revising the selection history; 4) ability to communicate, and to be integrated within an eLearning platform, such as ILIAS or MOODLE.

A demo version of the Open Key Player running a key on woody plants of Romania can be found at the URL specified at [9]. The application requires the input of two parameters: the name of the database which contains the identification key, and the code of the specific key in the database (since the same database can store multiple identification keys).

The user interface is divided into two panels (Fig. 2). The left panel contains a "Reset" button, which starts the identification process from the beginning, and reports the number of remaining species and the history of the identification steps ("Selected Features"). The right panel contains the current choices of the identification process, which can be illustrated by images. The counter of remaining species is updated, and the last step is added to the "Selected Features" list at each step of the identification process.

When one of the choices leads to a result (Fig. 3), the scientific name of the taxon appears in brackets. The information related to the identified taxon is displayed in a pop-up window at the end of the identification process. It can also contain an image gallery, when this gallery is available in the database.

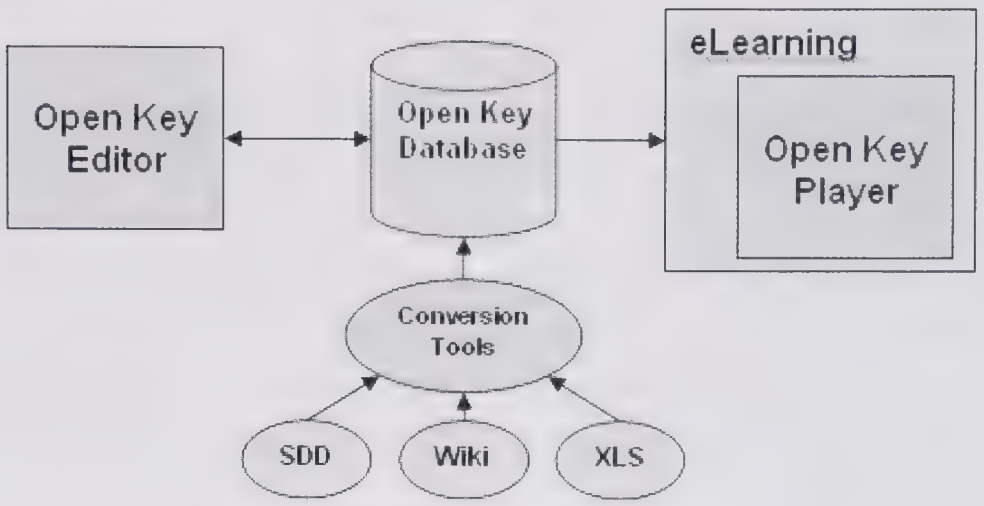


Fig. 1 – The architecture of the *KeyToNature* Open Key Standard (Editor, Player, Conversion tools).

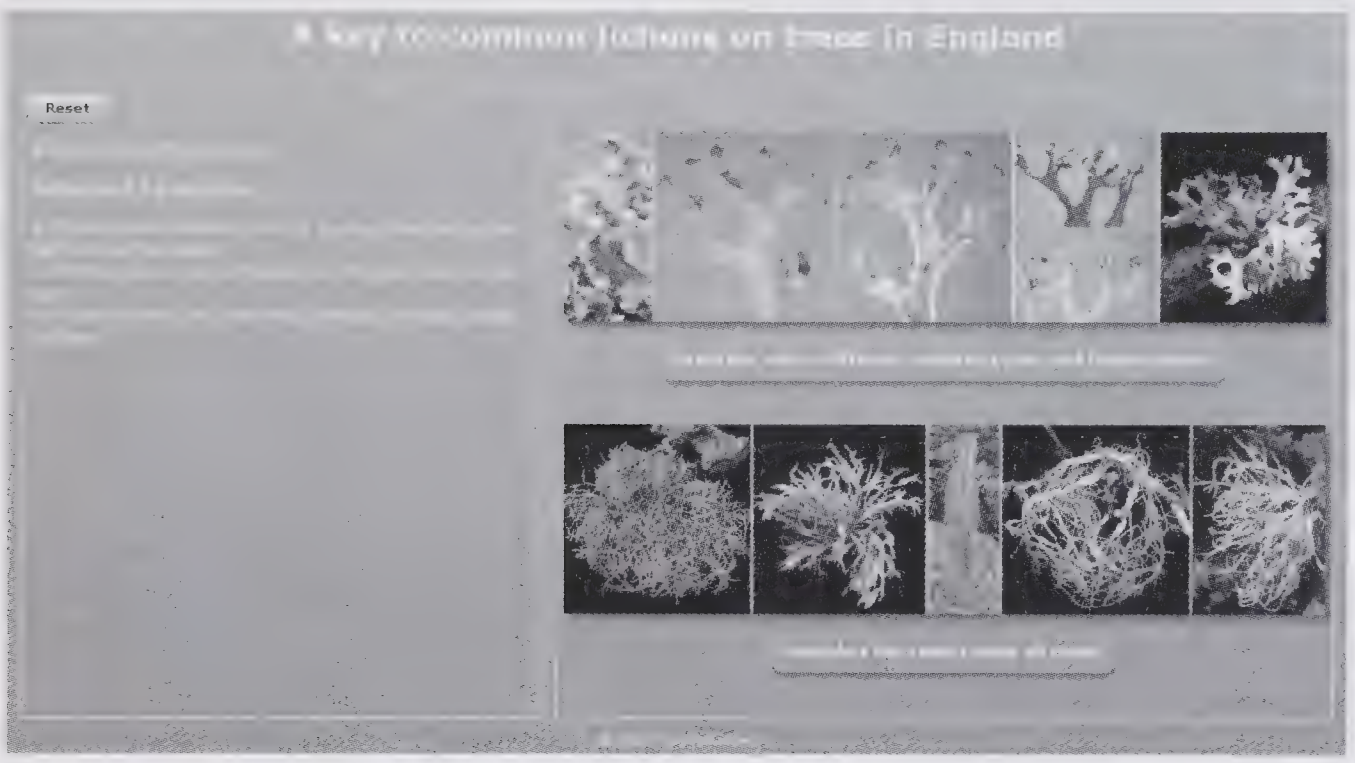


Fig. 2 – The interface for the Open Key Player (Left: selection history, Right: selection options).

Usnea cornuta Körb.
Are you satisfied with the identification result?
Yes
No
Common name: none.
Diagnostic characters: Thallus green, tufted, ± erect, main branches slightly inflated, with branches constricted where they join the main stem, which is stretchy when pulled. Holdfast pale at base. Soredia in irregular patches on the surface of branches amongst isidia-like structures. Fruits not known. Spot tests K+ yellow turning red, P+ yellow-orange.
Photobiont: green chlorococcoid alga.
Distinguished from: U. subfloridana, which is black at the base, where it is attached to the bark, with a tough main stem that does not stretch when pulled.
Ecology: widespread on twigs, branches and trunks and now appearing in previously acid rain polluted habitats.
Distribution: Widespread in W and SW of the UK.

Fig. 3 – The panel with the final result of the identification process.

4 User Tracking

One of the most interesting features of the OKP is the possibility of integrating it in an eLearning environment, thus providing user-tracking. The application can communicate to the eLearning environment each interaction made by the user. This feature allows the creators of identification keys to access interesting statistics on users behaviour, and to use them to improve the keys. The Open Key Player has been successfully integrated in ILIAS, and was tested in several Romanian high schools, giving back valuable information about the key of woody plants of Romania.

5 Conclusions

The OKP has proven to be a valuable asset for the Open Key Standard of *KeyToNature*. It provides a modern, user-friendly interface, which it can be integrated into eLearning environments, providing user-tracking statistics. The tests carried out in the Romanian high schools showed that the application is an efficient interactive tool, and that it could be an important component in teaching and learning biodiversity.

Acknowledgement

This paper has been supported by the project *KeyToNature* (www.keytonature.eu, ECP-2006-EDU-410019), in the framework of the *e*Content*plus* Programme.

References

[1] S. Martellos, "Multi-authored interactive identification keys: The FRIDA (FRiendly IDentificAtion) package" *Taxon,* vol. 59 (3), pp. 922-929, 2010.

[2] S. Martellos, E. van Spronsen, D. Seijts, N. Torrescasana Aloy, P. Schalk and P. L. Nimis, "Digital identification keys to organisms and user-generated content. The *KeyToNature* approach", *Proceedings of the IASK International Conference Teaching and Learning*, pp. 96-102, 2009.

[3] S. Martellos, E. van Spronsen, D. Seijts, N. Torrescasana Aloy, P. Schalk and P. L. Nimis, "User-generated content in the digital identification of organisms: the *KeyToNature* approach", *Int. J. Information and Operations Management Education,* vol. 3, 3, pp. 272-83, 2010.

[4] J. D. Herrington and E. Kim, *Getting started with Flex 3*, O'Reilly Media Inc., 2008.

[5] C. E. Brown, *The Essential Guide to Flex 3*, Apress, 2008.

[6] E. R. Harold, and W. S. Means, *XML in a nutshell*, O'Reilly Media Inc., 2004.

[7] E. T. Ray, *Learning XML*, O'Reilly Media Inc., 2003.

[8] D. J. Barrett, *MediaWiki*, O'Reilly Media Inc., 2008.

[9] Demo of OKE: http://octopus.utcluj.ro:56340/okp/openKeyPlayer.swf?db=oke&key=1, July 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 137-143.
ISBN 978-88-8303-295-0. EUT, 2010.

Improvement of identification keys by user-tracking

Gerd Schmidt, Mircea Giurgiu, Sónia Hetzner, Fred Neumann

Abstract — Identifications keys are indispensable tools for identifying organisms and understanding biodiversity. It is therefore advantageous to give a broad public access to these tools. Due to the fact that these tools are used by very different target groups – pupils, students, researchers, etc. – customisation and quality enhancement are crucial for effective and appealing identification keys. The evaluation of key software by inquiring users is always influenced by various, often subjective impressions and is usually restricted to a small number of them. The identification keys developed in the framework the *KeyToNature* project are available on electronic devices and allow registering the actions of single users during the identification process. By transferring user actions to an Learning Management System (LMS) like ILIAS OpenSource, where they can be correlated to user-data like experience in the field of identification, courses visited, age etc., it is not only possible to identify the problems a individual user did encounter and the mistakes he made, but also to conduct statistical evaluations to review and substantially improve the keys, to solve problems with single questions, and to optimise customisation to the different user groups.

Index Terms — identification keys, user-tracking, eLearning, Rich Internet Application, customisation, evaluation.

1 Introduction

Identification keys (id-keys) are indispensable tools to identify organisms, which is a basic process for understanding nature and preserving biodiversity. The European project *KeyToNature* aims at developing and improving identification keys for all media, including electronic devices such as smartphones and all types of computers. [1]. This form of electronic keys permits – as a fundamental new possibility – to record every action performed by users of a key, which is named here "tracking-feature". Every wrong alternative chosen by the user, every achieved species name, accepted as right or rejected as wrong, can be registered. Gathering these information in a Learning Management Software (LMS) [2], where it can be related to all user-

G. Schmidt, S. Hetzner and F. Neumann are with the FIM Innovation in Learning Institute, Nägelsbachstr. 25b in 91052 Erlangen. E-mail: gerd.schmidt@fim.uni-erlangen.de, sonia.hetzner@fim.uni-erlangen.de; fred.neumann@fim.uni-erlangen.de.
M. Giurgiu is with the Universitatea Tehnica din Cluj-Napoca, Dep. Telecomuncatii, Str. Baritiu Nr. 26, 400020 Cluj-Npoca. E-mail: Mircea.Giurgiu@com.utcluj.ro.

specific data (pre-knowledge, courses taken, age) enables a new benchmarking of identification software and keys.

Therefore an add-on for the LMS ILIAS [3] was developed, that consists of a *logging service* to record every action a user performs in an id-key, and an *evaluation service* that performs basic statistical analysis for continuous monitoring and quality enhancement.

The "*logging service*" in connection with the "*tracking feature*" inside the *Flex Player* [4] or the *Open Key Player* [5, 6] enables us to track user behaviour for continuous improvement of id-keys, to correlate the results with user groups: age, education, previous experiences etc., and to gather an enormous amount of information about how users use the key.

The "evaluation service" with filtering and exporting features offers automated-evaluation processing at low efforts, testing the eligibility of keys for user groups (e.g. age) by the filtering feature and exporting all information for further and more detailed analysis.

2 What we want to know exactly

The above described services aim at continuously providing updated user requirements for the further development and customisation of the id-keys. More concretely, information such as: How long users need to orient themselves in the key (especially true for multi-access id-keys)? How does the user navigate in the key (choose an alternative, undo a choice, and restart because of feeling lost)? How long does a user need to answer every single question? How much time does a user need for one identification run? Does she or he reach the correct result at the first attempt? Does she or he feel satisfied or certain to have correctly identified the organism? How long does she or he use the key at all?

Additionally it is also extremely interesting to analyse the influence of pre-knowledge and use of the id-key, as well as the specific pedagogical settings. Furthermore, it is most interesting to investigate further factors that influence the results, such as age, gender, school type, etc.

To answer these questions, it is necessary to get information about every single event within the usage of a key, and its duration. All data are gathered by the logging service, and a first automatic analysis is performed by the evaluation service that was first developed for ILIAS and in the Open Key Player (single-access keys) and the Flex Player (multi-access keys).

3 How does the process work

3.1 Integration Overview

The information flow in the collaboration between the LMS (ILIAS) and the key software is shown in Fig. 1.

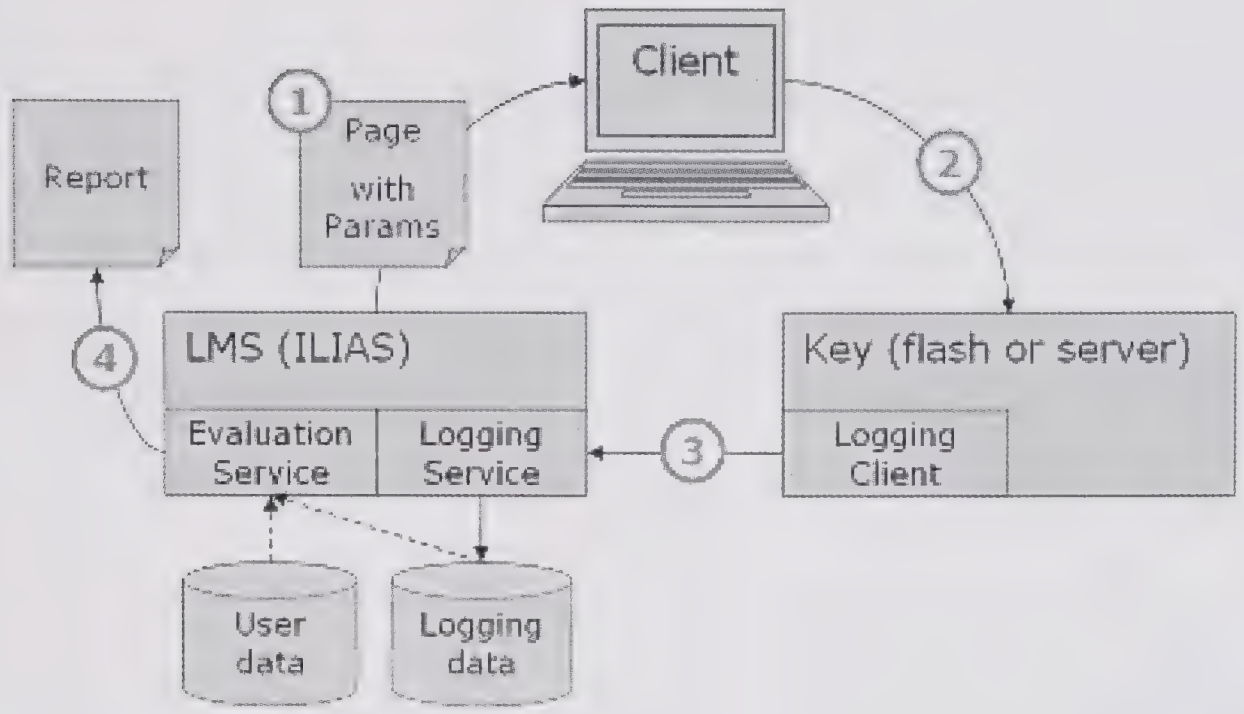


Fig. 1 – Information flow LMS - User - Key-Software - LMS - Outside.

3.2 Data transmitted to the Key (Step 1 and 2)

The information transmitted to the Open Key Player software consists of the name of database to be used by the Player, a reference-id to specify the ILIAS object representing the key, and a user-specific session-id to correlate the tracking of data to the individual user when they are sent back to the LMS.

3.3 Logging of User Interactions (Step 3)

The key sends back every user action to the LMS by appending them to the URL of the logging service, together with the reference-id and the session-id. The log entry is written by ILIAS only if the session_id is valid. The logging service is responsible to add a time stamp to every action recorded.

3.4 Evaluation Service (Step 4)

The evaluation service inside the LMS performs the tasks of filtering and exporting the log events by user profile data (age range, education, precognition, e.g., lessons or courses attended before the key was used). It is also able to perform a first analysis of data such as a count of the total number of key sessions (how often the key was used at all), the average time of a key session (is the key exciting or boring to the users?) or the average time to answer a question (is it simple or difficult?). It can also count how often an answer was revised, etc.

4 Logging events provided by the key respectively by the key-player software

The information sent back by the key and logged by the LMS was designed to serve a broad list of types of keys, and should be open for future developments. Therefore it was decided to log four parameters that can be freely filled with text-

strings or numeric values.

4.1 Example for tracking a multi-access key

Multi-access keys were the first type for which user-tracking was established. In this type of keys, users can select the question they want to answer first, so that they need some time to orient themselves in the key. At the end of the identification they can decide if they are confident with the result or if they want to restart the determination process. With the sorting and filtering feature of the logging service the log-events can be filtered and exported for further evaluation.

USER	TIME	EVENT TYPE	EVENT SUBTYPE	VA	EVENT TEXT
user, root	2009-09-25 11:11:36	Flex Key Player Start	Application First Load	0	
user, root	2009-09-25 11:12:16	Identification Step	Character Selected	0	Does it have legs?
user, root	2009-09-25 11:12:20	Identification Step	State Selected	0	Yes, it has legs
user, root	2009-09-25 11:12:20	Identification Step	First Step Taken	0	Character: Heeft het dier poten?; State: Ja, di
user, root	2009-09-25 11:13:00	Identification Step	Character Selected	0	What is the colour on the back?
user, root	2009-09-25 11:13:05	Identification Step	State Selected	0	Black and yellow
user, root	2009-09-25 11:13:05	Identification Step	New Step Taken	0	Character: Wat voor kleur kun je zien op de ru
user, root	2009-09-25 11:13:05	Identification End	Success	0	Identification ended with result: Salamandra s
user, root	2009-09-25 11:13:11	Identification End	User is satisfied with result	0	Identification ended with result: Salamandra s
user, root	2009-09-25 11:13:21	Application Reset	Reset	0	Reset caused either by user pushing Reset b
user, root	2009-09-25 11:13:21	Flex Key Player Start	Application First Load	0	
user, root	2009-09-25 11:13:26	Identification Step	Character Selected	0	Does it have a tail?
user, root	2009-09-25 11:13:31	Identification Step	State Selected	0	Yes, it has a tail
user, root	2009-09-25 11:13:31	Identification Step	First Step Taken	0	Character: Heeft het dier een staart?; State: J
Pupil, 1	2009-09-25 11:20:30	Flex Key Player Start	Application First Load	0	
Pupil, 1	2009-09-25 11:28:43	Identification Step	Character Selected	0	Does it have legs?
Pupil, 1	2009-09-25 11:28:45	Identification Step	First Step Taken	0	Character: Heeft het dier poten?; State: Ja, di
Pupil, 1	2009-09-25 11:28:45	Identification Step	State Selected	0	Yes, it has legs
Pupil, 1	2009-09-25 11:28:55	Identification Step	Character Selected	0	Does it have a carapace?
Pupil, 1	2009-09-25 11:28:57	Identification Step	New Step Taken	0	Character: Heeft het dier een schild op zijn ru
Pupil, 1	2009-09-25 11:28:57	Identification Step	State Selected	0	Yes, it has a carapace
Pupil, 1	2009-09-25 11:28:57	Identification End	Success	0	Identification ended with result: Trachemys sc
Pupil, 1	2009-09-25 11:29:06	Identification End	User is NOT satisfied with result	0	Identification ended with result: Trachemys sc
Pupil, 1	2009-09-25 11:29:14	Application Reset	Reset	0	Reset caused either by user pushing Reset b
Pupil, 1	2009-09-25 11:29:14	Flex Key Player Start	Application First Load	0	
Pupil, 1	2009-09-25 11:29:17	Identification Step	Character Selected	0	Does it have legs?
Pupil, 1	2009-09-25 11:29:22	Identification Step	State Selected	0	Yes, it has legs
Pupil, 1	2009-09-25 11:29:22	Identification Step	First Step Taken	0	Character: Heeft het dier poten?; State: Ja, di
Pupil, 1	2009-09-25 11:29:36	Identification Step	Character Selected	0	What is the colour on the back?
Pupil, 1	2009-09-25 11:29:39	Identification Step	New Step Taken	0	Character: Wat voor kleur kun je zien op de ru
Pupil, 1	2009-09-25 11:29:39	Identification End	Success	0	Identification ended with result: Salamandra s
Pupil, 1	2009-09-25 11:29:42	Identification End	User is satisfied with result	0	Identification ended with result: Salamandra s

Fig. 2 – Tracking example for a multi-access key
(the lines have been coloured manually).

Data are sorted by user and time, so that the actions of single users within the key can be easily observed. For example, pupil 1 performs an identification step but is “NOT satisfied with the result” reached in the first step of the identification. Thus, he does an “Application Reset” and finishes with the correct result. Statistical analysis of large amounts of tracking data can identify certain questions that often lead to wrong results, or are re-visited many times. Accordingly, these specific questions are improved and closely evaluated in further testing events.

4.2 Example for the tracking of a single-access key

The following table was extracted from the log file of a determination of trees and shrubs in a Romanian school. The “History revision” (marked in yellow) represents selections that have been revised. The option number 198 was independently re-visited by more pupils, which indicates that this alternative is not clear enough.

User	Time stamp	event_type	event_subtype	event_integer question number	event_text selected alternative
keytopm14	09:50:04	Identification Step	Character Selected	182	Plante cu diferite caractere
keytopm14	09:50:08	Identification Step	Character Selected	185	Plante nepungente, f?r? spini
keytopm14	09:50:20	Identification Step	Character Selected	198	Frunze cu contur mai mult sau mai pu?in t
keytopm14	09:50:25	Hystory Revision	Character Selected for Revsion	198	Frunze cu contur mai mult sau mai pu?in t
keytopm14	09:50:27	Identification Step	Character Selected	198	Frunzele nu au contur triunghiular sau rom
keytopm14	09:50:36	Identification Step	Character Selected	205	Frunze truncate, cordate sau hastate la ba
keytopm14	09:50:39	Identification Step	Character Selected	206	Marginea frunzei din?at?
keytopm14	09:50:41	Identification Step	Character Selected	207	Frunze verzi pe ambele fe?e, nu sunt puter
keytopm14	09:50:50	Identification Step	Character Selected	209	Fruct diferit
keytopm14	09:50:53	Hystory Revision	Character Selected for Revsion	209	Fruct diferit
keytopm14	09:50:54	Identification Step	Character Selected	209	Fruct uscat ?i dezvoltat pe pedicel lung, cu
keytopm14	09:50:59	Identification Step	Character Selected	210	Fa?a inferioar? a frunzei cu peri ruginii la b
keytopm14	09:50:59	Identification End	Success	0	Identification ended with result: Tilia cordat
keytopm14	09:51:09	Identification End	User is satisfied with result	0	Identification ended with result: Tilia cordat
keytopm15	09:35:17	Application Start	Start	0	Identification started
keytopm15	09:35:34	Identification Step	Character Selected	1	Plante înr?d?cinate în sol
keytopm15	09:36:13	Identification Step	Character Selected	185	Plante nepungente, f?r? spini
keytopm15	09:36:16	Identification Step	Character Selected	198	Frunzele nu au contur triunghiular sau rom
keytopm15	09:36:23	Identification Step	Character Selected	205	Frunze rotunjite sau îngustate la baz?
keytopm15	09:36:28	Hystory Revision	Character Selected for Revsion	198	Frunzele nu au contur triunghiular sau rom
keytopm15	09:36:30	Identification Step	Character Selected	198	Frunze cu contur mai mult sau mai pu?in t
keytopm15	09:36:42	Hystory Revision	Character Selected for Revsion	185	Plante nepungente, f?r? spini
keytopm15	09:36:51	Identification Step	Character Selected	185	Plante nepungente, f?r? spini
keytopm15	09:36:56	Identification Step	Character Selected	198	Frunzele nu au contur triunghiular sau rom
keytopm15	09:37:42	Identification Step	Character Selected	209	Fruct uscat ?i dezvoltat pe pedicel lung, cu
keytopm15	09:37:49	Identification Step	Character Selected	210	Fa?a inferioar? a frunzelor cu peri albicio?i
keytopm15	09:37:52	Hystory Revision	Character Selected for Revsion	210	Fa?a inferioar? a frunzelor cu peri albicio?i
keytopm15	09:38:18	Identification Step	Character Selected	210	Fa?a inferioar? a frunzei cu peri ruginii la b
keytopm15	09:38:18	Identification End	Success	0	Identification ended with result: Tilia cordat
keytopm15	09:39:27	Identification End	User is satisfied with result	0	Identification ended with result: Tilia cordat

Fig. 3 – Tracking example for a multi-access key (Grey lines represent tracking events that have been removed to restrict the length of the table).

5 Feedback from the user tracking to the key builders by analysis of the logfile

Without any explicit analysis, simply by sorting and filtering functions, the logfile provides insight into how single users move in the key, how many times they reached a result, rated it as right or wrong. It also provides a list of the species they found with the key or how many organisms they tried to identify. The logfile of the user tracking can be statistically analysed, basically in the LMS with the evaluation service itself, and in a more extended and detailed way when exported as an Excel-sheet for further analysis.

The following calculations may be used as feedback to improve the key and are directly provided by the logging service in the LMS:

1. Key starts / Results assumed as correct or wrong
2. Time to select a question (multi-access keys only)
3. Time to answer a question
4. Time to successfully identify a species
5. Most selected questions (multi-access keys only)

To get additional, statistical information, the log-file can be exported for more detailed analysing.

school 1	Species given	Species found	Identification ended	Not satisfied	Satisfied
	Abies alba	Abies alba Mill.	7		7
		Abies nordmanniana (Steven) Spa	3		2
		Actinidia chinensis Planch.	1		
	Aesculus hippocastanum	Aesculus hippocastanum L.	7		6
		Bruckenthalia spiculifolia (Salisb.)	1		1
	Buxus sempervirens	Buxus sempervirens L.	1		
		Campsis radicans (L.) Seem.	1		1
		Cupressus sempervirens L.	1		
	Ginkgo biloba	Ginkgo biloba L.	10		10
		Hippophaë rhamnoides L.	2	1	
		Juniperus communis L.	1		
		Juniperus communis L. subsp. na	4		1
		Loranthus europaeus Jacq.	3		2
		Maclura pomifera (Raf.) C.K. Schr	2	1	1
		Myricaria germanica (L.) Desv.	1	1	
	Picea abies	Picea abies (L.) H. Karst.	2		1
		Picea pungens Engelm.	1		
		Populus alba L.	1		
	Populus nigra	Populus nigra L.	6		5
		Satureja montana L.	1		
		Taxodium distichum (L.) Richard	1		
		Taxus baccata L.	2		1
	Thuja occidentalis	Thuja occidentalis L.	6	1	5
	Thuja orientalis	Thuja orientalis L.	7		5
	Viscum album	Viscum album L.	11		10

Pupils:	18
Identifcation total	83
Identifications minus "not satisfied"	4
Identifications affirmed with "satisfied with result"	58
Events	841
Events / Pupils	47

Fig. 4 – Log-File of class A after second level analysis. “Species found” refers to the result the identification.

6 Reliability and Performance

The testing events that took place in a Romanian school have been conducted with two classes, concurrently using the Open Key Player software that was fed by a Dutch database, sending the log data to a LMS hosted in Germany. Nearly 4000 datasets were recorded without problems and without significant load to the servers. The logging service is estimated to be reliable and scalable, and able to gather large amounts of information about the usability of keys in Europe.

7 Conclusions

The combination of identification tools with tracking features and a logging and evaluation service in an LMS can give objective information about user behaviour and the quality of identification keys. The filtering features allow differentiating the appropriateness of keys for different user groups and enable the authors of keys to improve them over time, adapting them to the needs of different target groups. The application is robust and reliable, and can handle a high number of concurrent users. As the actual evaluation of a single key can be conducted by hundreds of LMS-Systems, the next step will be to hand back user-specific quality indicators to the key or the key-database by means of the key-player software.

Acknowledgement

This paper has been supported by the project *KeyToNature* (www.keytonature.eu, ECP-2006-EDU-410019), in the framework of the *e*Content*plus* Program.

References

[1] L. M. Systems, http://www.e-teaching.org/technik/distribution/lernmanagementsysteme, 2010.

[2] ILIAS: http://www.ilias.de/docu/, 2010.

[3] S. Martellos, E. van Spronsen, D. Seijts, N. Torrescasana Aloy, P. Schalk and P. L. Nimis, "Digital identification keys to organisms and user-generated content. The *KeyToNature* approach", *Proceedings of the IASK International Conference Teaching and Learning*, pp. 96-102, 2009.

[4] M. Giurgiu, G. Hagedorn and A. Homodi, "IBIS-ID, an Adobe FLEX based identification tool for SDD-encoded multi-access keys". *Proc of TDWG 2009*, 9-13 November 2009, Montpellier, p. 90, 2009.

[5] M. Giurgiu, A. Homodi, E. van Spronsen, S. Martellos and P. L. Nimis, "Open Key Player: A new approach for online interaction and user tracking in identification keys". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 133-136, 2010.

[6] Demo version of Open Key Player, http://octopus.utcluj.ro:56340/okp/openKeyPlayer.swf?db=oke&key=1, July 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 145-150.
ISBN 978-88-8303-295-0. EUT, 2010.

ARIES: an expert system supporting legislative tasks. Identifying animal materials using the Linnaeus II software

Leo W.D. van Raamsdonk

Abstract — Bovine Spongiform Encephalopathy (BSE, Mad cow disease) is generally considered to be caused by recycling animal by-products as ingredient in animal, especially ruminant, feed. Feed bans were enforced to minimize the risk on infections, and monitoring programs are effectuated for controlling the ban. The only official detection method is visual (microscopic) examination of the presence of primarily bone fragments, but muscle fibres, hairs, feather filaments, and fish bones are targeted as well. An expert system is developed called ARIES (Animal Remains Identification and Evaluation System) with identification tools for those fragments. ARIES provides information on procedures, legislation and background documentation, and descriptions of prohibited materials as well as of a range of confusing particles. The system allows documentation of the results of evaluations for future reference and is considered a good platform for training and education.

Index Terms — ARIES, BSE, identification, Linnaeus II software, mad cow disease.

1 Introduction

It is generally supposed that the most likely route of infection of cattle with Bovine Spongiform Encephalopathy (BSE, mad cow disease) is by means of feeds containing low levels of processed animal proteins (PAPs). Because of this likely route of infection, feed bans were enforced, primarily for ruminant feeds, but later on extended to all feeds for farmed animals. The goal of legislation is the enforcement of the "species-to-species" ban which prohibits only the feeding of animal specific proteins to its own species [1]. However, such a species-to-species ban needs to be supported by species-specific identification methods.

In the European Union microscopic evaluation is currently the only accepted

L. W. D. van Raamsdonk is with RIKILT – Institute of food safety, P.O. Box 230, 6700 AE Wageningen, the Netherlands. E-mail: Leo.vanRaamsdonk@wur.nl.

method for the detection and characterization of PAPs in feeds [2], [3]. This method is predominantly focused on the presence and characteristics of bone fragments, although other structures, e.g. muscle fibres, may provide circumstantial evidence of the respective animal types. Recent developments are the identification of bone fragments at the level of classes (mammal vs. bird vs. fish), supported by image analysis of bone characteristics [4].

Identification of bone fragments is based on a series of characteristics, ranging from the shape and transparency of the bone fragments, to histological differences between the different species groups. Information on as much characteristics as possible should be collected when a few bone particles are detected in a feed sample in order to get an acceptable reliability of a presumed identification. Nevertheless, in many cases it is impossible to get a 100% match between all collected information and the profile of one of the targeted species. Consequently, uncertainty analysis is one of the necessary aspects of a scientifically sound evaluation of microscopic analyses.

2 Expert system ARIES

2.1 Development

An expert system has been developed to document every aspect of the detection and identification of prohibited animal materials in feed. The software package Linnaeus II (ETI, Amsterdam, [5]) has been chosen as platform. A first stand-alone version of the expert system was released in 2004 with the name ARIES (Animal Remains Identification and Evaluation System), as a product of an EU funded project ("STRATFEED", [6]). ARIES provides a series of modules, including an Introduction, full documentation on protocol and methods, an overview of prohibited and confusing materials (e.g. plant hairs, minerals of animal origin), several identification trees and matching procedures, a glossary, an overview of legislation and extensive documentation from literature and internet links [7].

A version 2.0 has been developed as a web-based application in the framework of the EU funded project "SAFEED-PAP" [8]. The release of ARIES v2.0 was achieved in 2010 [9].

2.2 Validation of the use of ARIES

ARIES has been used in a validation study of the microscopic method for the detection of animal proteins in feed according to the official method [10]. In this study, 25 laboratories investigated a set of 24 blind samples, partly adulterated with several types of animal proteins in seven different treatments including a control (blank). Thirteen of these laboratories used ARIES for support of their detection and identification of the materials, twelve did not. All participants were asked to report the presence of fish meal, material of terrestrial animals, and if the latter was present, to indicate whether it was mammalian or avian material.

In all cases a "presence", "absence" or "no result" could be reported. The results were summarised in accuracy values: the number of correct results divided by the total number of reports (excluding the number of no results).

This presentation will focus on the results for the proper detection of mammalian material in the appropriate treatments and the presence of confusing ingredients (fish meal). Therefore, the results for four different treatments (sample types) were shown:

1. blank feed
2. feed with 5% of fish meal
3. feed with 0.1% of MBM and 5% of fish
4. feed with 0,1 % of MBM

The overall scores for these treatments and parameters, not stratified for the use of ARIES, are presented in Tab. 1. The detection at the highest classification level (fish material, and terrestrial animal material) poses in general no problem. The detection of terrestrial animal material in the presence of fish material (0.768) should be improved.

A considerable number of laboratories felt insecure for the identification of specifically mammalian material, as is shown by the high number of "no results" in Tab. 1.

	n	AC		
Material		terrestrial	mammalian	fish
blank	100	0.908 (2)	0.933 (7)	0.880 (0)
fish 5%	100	0.857 (2)	0.919 (10)	0.990 (1)
MBM 0.1% + fish 5%	100	0.768 (1)	0.639 (27)	0.970 (0)
MBM 0,1 %	75	0.987 (0)	0.896 (27)	0.920 (0)

Tab. 1 – Basic results expressed in accuracy for the detection of different types of animal proteins in four differently contaminated feeds. Number of "No results" in brackets. n: total number of observations.

	AC with ARIES		AC without ARIES	
Material	n	mammalian	n	mammalian
blank	52	0.935 (2)	48	0.930 (5)
fish 5%	52	0.891 (2)	48	0.950 (8)
MBM 0.1% + fish 5%	52	0.702 (4)	48	0.520 (23)
MBM 0.1 %	39	0.857 (4)	36	1.0 (23)

Tab. 2 – Basic results expressed in accuracy for the detection of mammalian material in four differently contaminated feeds split for the use of the ARIES system. Number of "No results" in brackets. n: total number of observations.

	AC' with ARIES		AC' without ARIES	
Material	n	mammalian	n	mammalian
blank	52	0.896	48	0.833
fish 5%	52	0.854	48	0.792
MBM 0.1% + fish 5%	52	0.647	48	0.271
MBM 0.1 %	39	0.769	36	0.361

Tab. 3 – Recalculated results expressed in adjusted accuracy for the detection of mammalian material in four differently contaminated feeds divided in two groups of users and non-users of the ARIES system. n: total number of observations.

There is no significant difference in the performance of the laboratories that used or did not use ARIES for the detection of terrestrial animal material in general, and of fish material. However, for the identification of mammalian material low numbers of "no result" were reported for the group of ARIES users compared to the non-users (Tab. 2). The ARIES users apparently made use of the system as a sufficient source of expertise that allowed them to draw a (correct) conclusion in most cases.

The accuracy values for non-users are biased by the high number of "no results". Therefore corrected accuracy values were calculated (Tab. 3). From these values it can be concluded that ARIES offers a considerable support to the detection and identification of mammalian material.

2.3 Evaluation

The participants of the validation study showed, by means of investigating a training sample set prior to the proper study, to have a sufficient level of expertise for the type of material that was included in the validation study. This starting situation explains why there is no significant difference between the group of users and of non-users for the detection of terrestrial animal proteins in general, and of fish meal. It should be emphasized, however, that ARIES provides a much larger range of data and information than the subset that was necessary for performing well in the validation study. As an example, unintentional contamination of feeds by rodents can be identified by using the information on different hair types as included in ARIES.

2.4 Application

A European network of more than 100 official control laboratories monitors the ban on animal materials in animal feed. The backbone to this monitoring is provided by an EU network of National Reference Laboratories, one for each Member State, coordinated by the European Union Reference Laboratory. Approximately 50 of these laboratories applied for a subscription to the stand-alone version 1.0. ARIES is used frequently because it offers assistance in the application of the methodology and the identification of the findings. The system allows documentation of the results of the evaluation for future reference and is

considered a good platform for training and education.

3 Discussion

The application of an expert system for support of the detection of prohibited animal proteins in feed is an advantage in several ways:

- Support of daily routine analyses by providing procedural information as well as the evaluation of observations..
- A formalised evaluation of observations as documentation for later reference.
- Training system and a platform for knowledge transition.

In the case of ARIES, the system is designed to be helpful for the experienced scientists, as well as for training and e-learning of less experienced microscopists.

The development of new information was necessary for a new version 2.0. The project SAFEED-PAP [8] provided a sufficient amount of data for a principle improvement and fine tuning. The choice to develop a web application allows to update the system whenever new information needs to included. A web based application also implies that a sustainable support should be maintained. It is the intention to give users access to ARIES 2.0 with a username and password based on a reasonable annual fee. In this way maintenance can be assured without having a commercial exploitation.

The performance of the microscopic method as illustrated by the accuracy indices of Tab. 1 reflect the situation in 2004. After that, improvements have been achieved, and in a period of five years an accuracy of 0.98 was established in a blind test for European laboratories for the detection of 0.1% of MBM in the presence of 5% of fish material [3], [11]. ARIES includes descriptions of prohibited materials and a range of confusing plant materials in order to minimise false positive detections. By providing this information, ARIES is one of the tools for maintaining this high level of performance.

Acknowledgements

This work was supported by the European Commission in the framework of the European Project SAFEED-PAP (FOOD-CT-2006-036221), "Detection of presence of species-specific processed animal proteins in animal feed", funded under the 6th EC FP, DG RTD.

References

[1] European Union, "Regulation (EC) No 1774/2002 laying down health rules concerning animal by-products not intended for human consumption". *Official Journal of the European Communities*, 10.10.2002, L 273, pp.1-95, 2002.

[2] European Commission, "Commission Regulation (EC) No 152/2009 of 27 January 2009 laying down the methods of sampling and analysis for the official control of feed". *Official Journal of the European Communities* L 54, 26.2.2009, pp. 1–130, 2009.

[3] L. W. D. van Raamsdonk, C. von Holst, V. Baeten, G. Berben, A. Boix and J. de Jong, "New developments in the detection of animal proteins in feeds". *Feed Science and Technology,* vol. 133, pp. 63-83, 2007.

[4] A. Campagnoli, C. Paltanin, G. Savoini, A. Baldi and L. Pinotti, "Combining microscopic methods and computer image analysis for lacunae morphometric measurements in poultry

and mammal by-products characterization". *Biotechnol. Agron. Soc. Environ.,* vol. 13(S), pp. 25-28, 2009.
[5] ETI bioinformatics. *Linnaeus II software package*. http://www.eti.uva.nl/, 2010.
[6] Project website: http://stratfeed.cra.wallonie.be/, 2010.
[7] Vermeulen Ph., V. Baeten, P. Dardenne, L. W. D. van Raamsdonk, R. Oger, A.S. Monjoie and M. Martinez, "Development of a website and an information system for an EU R&D project: the example of the STRATFEED project". *Biotechnol. Agron. Soc. Environ*., vol. 7, pp. 161-169, 2003.
[8] Project website: http://safeedpap.feedsafety.org/, 2010.
[9] RIKILT Institute of food safety. *ARIES, Animal Remains Identification and Evaluation System*, http://aries.eti.uva.nl/, 2010.
[10] C. von Holst, L. W. D. van Raamsdonk, V. Baeten, S. Strathmann and A. Boix, "The validation of the microscopic method selected in the Stratfeed project for detecting processed animal proteins". *Stratfeed, Strategies and methods to detect and quantify mammalian tissues in feedingstuffs, chapter 7*. Office for Official Publication of the European Communities, Luxembourg, 2005.
[11] L. W. D. van Raamsdonk, W. Hekman, J. M. Vliege, V. Pinckaers and S.M. van Ruth, "Animal proteins in feed. IAG ring test 2009". *Report 2009.017, RIKILT*, Wageningen, 34 pp., 2009.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 151-156.
ISBN 978-88-8303-295-0. EUT, 2010.

An integrated system for producing user-specific keys on demand: an application to Italian lichens

Juri Nascimbene, Stefano Martellos, Pier Luigi Nimis

Abstract — The identification of lichens is important in several applied fields, such as the biological monitoring of air pollution and the restoration of open-air stone monuments. This often creates relevant problems for non-specialists and technicians which are in charge of routinely applying lichen monitoring techniques. The coupling of a complex information system (ITALIC), together with a new software which can automatically produce identification keys for any subset of species included in a database (FRIDA), is an innovative approach in the field of identifying biodiversity. ITALIC is able to produce a list of species which potentially occur under a set of ecological and distributional conditions specified by the user. The list is automatically transferred to FRIDA, which generates a user-oriented interactive identification key limited to the species present in the "virtual habitat" created by the user. The new system has relevant applications, since it effectively supports the technical personnel of Environmental Agencies, Nature Parks, Cultural Heritage Conservation Agencies involved in lichen monitoring throughout the Country.

Index Terms — interactive keys, biomonitoring, restoration of monuments, lichens, identification.

1 Introduction

Lichens are a diverse and species-rich group of fungi living in close symbiotic relationship with algae or cyanobacteria. Many of them form crustose thalli whose identification usually requires the analysis of micro-morphological or chemical characters. However, the identification of lichens is important in several applied fields, such as the biological monitoring of air pollution and the restoration of open-air stone monuments.

Epiphytic lichens are highly sensitive to environmental changes and air pollution, and they are among the most widely used biomonitors in terrestrial environments [1]. Epi- and endolithic lichens are important in the biodeterioration

The authors are with the University of Trieste, Department of Life Sciences, via Giorgieri, 10 – I 34127 Trieste. E-mail: junasc@libero.it, martelst@units.it, nimis@units.it.

of stone monuments [2], [3]. In Italy their identification is required for restoration programs, both in the planning phase and for monitoring the effectiveness of restoration practices [4]. The identification at species level often creates relevant problems to non-specialists and technicians which are in charge of routinely applying lichen monitoring techniques.

In Italy, the popularization of lichen biomonitoring was supported and encouraged by rendering available online information on lichens provided by researchers. The first relevant contribution was the Italian checklist by Nimis [5], which was the basis for the creation of ITALIC, the Information System on Italian Lichens [6], that is freely available online since 2003. ITALIC provides support to the identification process by offering the possibility to compare a given specimen with high-resolution pictures and ecological-distributional information. However, while helpful, this simple comparison is far from being enough for this difficult groups of organisms, especially for a layperson.

In this paper we present a system which is able to produce ”keys on demand”, by coupling lists of species for “virtual habitats” created by users in ITALIC with a software that automatically generates identification keys to the species in those lists.

2 Basic information contained in ITALIC

For each lichen species which occurs in Italy (ca. 2350 in total), ITALIC provides: 1) updated literature on the specimens recorded from each of the 20 administrative regions of the country, 2) a predictive map depicting its potential distribution in 9 bioclimatic regions of Italy, 3) high-resolution pictures, and 4) information on biological traits, ecological requirements, and commonness/rarity. Biological traits include: growth forms (crustose, narrowly-lobed foliose, large lobed foliose, fruticose, fruticose-filamentose, squamulose), photobiont type (Cyanobacteria, Trentepohlia, Chlorococcoid green algae), and reproductive strategy (sexual vs. asexual by soredia, isidia or thallus fragmentation). Ecological requirements are summarised by 4 ecological indicator values: pH, air humidity, solar irradiation and eutrophication. Each ecological indicator value is expressed on a 1-5 ordinal scale. An additional indicator, expressed on a 0-3 ordinal scale, describes the tolerance of a species to human disturbance (poleophoby or hemeroby). Information on altitudinal range and substrates (epiphytic, lignicolous, foliicolous, saxicolous, and terricolous) is given as well.

The commonness/rarity of a species is expressed by a scale with 8 classes, ranging from extremely rare to extremely common in each of the main bioclimatic areas of the Country. This parameter was assessed on the basis of: a) number of samples in the TSB lichen herbarium, b) number of literature records, and, c) expert judgement. The ‘extremely rare’ status is given only to taxa known from less than 5 localities in Italy, or to those which were not mentioned in the literature in the last fifty years. Recently-described or dubious taxa are excluded from this category. The species in the ‘extremely rare’ class are likely candidates for the ‘threatened species’ category of the IUCN criteria [7], and therefore important for conservation purposes.

3 Automatic generation of species lists for "virtual habitats"

By combining several morphological, ecological and distributional parameters, ITALIC permits to elaborate complex queries for reconstructing "virtual habitats" in different parts of the country, the output being in the form of lists of species which are most likely to occur under the specified conditions. The predictivity of these virtual lists was tested by Nimis & Martellos [8] using a multivariate analysis of a matrix of "virtual" and real relevés of epiphytic lichen vegetation. Virtual relevés were obtained by selecting administrative region, altitudinal belt, substrate, and appropriate values for each ecological indicator value according to the main features of the targeted habitats. The results showed that the "virtual relevés" are highly predictive models, indicating that ITALIC could be consistently used for generating a large number of lists of lichens potentially occurring under conditions specified by the user. In the framework of the "Carta della Natura" project, promoted by the Italian Institute for Environmental Protection and Research (ISPRA), we recently applied this approach for providing potential lists of lichens for each of the ca. 160 CORINE-Biotopes habitats inventoried for Italy at the 1:50.000 scale [9]. The lists were obtained by combining the most important parameters describing the distribution and the ecology of Italian lichens: regional distribution, bioclimatic region, substrate type, ecological indicator values. In a few cases, additional parameters, such as commonness/rarity and tolerance to human disturbance, were used as well. A qualitative evaluation of the predictivity of these lists was assessed by checking real species lists available for well-studied habitats such as coniferous alpine forests [10], [11]: a good correspondence between the two datasets was found. Despite the fact that more quantitative evaluations on a large dataset would be welcome to statistically confirm previous results, this experience supports the practical utility of using the automatic generation of species lists for applied purposes.

4 FRIDA: from a morpho-anatomical database to interactive identification keys

In the last decade, our research team has developed original software for automatically generating interactive identification tools. A first phase of the research was conducted in the framework of the national project *Dryades,* and then continued in the European project *KeyToNature*. The most important software is FRIDA [12], a package which permits to generate interactive identification keys from a database of morpho-anatomical characters, starting from any list of species. The huge floristic information contained in several local checklists or vegetation studies can thus be easily used for generating new and original identification keys. For example, we used the detailed floristic information provided by Poldini [13] for the area of the natural reserve of the Val Rosandra (Trieste, NE Italy), and by Festi & Prosser [14] for the Paneveggio-Pale di San Martino Natural Park (Trento, NE Italy), to generate digital keys to the 988 and 1451 species of vascular plants known to occur in these two areas, respectively. In the last years, the continuous development of FRIDA, and of other original

software produced in the framework of *Dryades* and *KeyToNature,* permitted to greatly improve both the morpho-anatomical databases and the identification keys, which are now available in different and user-friendly layouts, such as those running on iPhones and other portable devices.

5 A STEP FURTHER: THE COMBINED USE OF ITALIC AND FRIDA FOR GENERATING INTERACTIVE IDENTIFICATION KEYS ON DEMAND

Here we provide an example of the high informative and applicative potential which derives from the combined use of an ecological-distributional database (ITALIC) with a morpho-anatomical database (FRIDA). The great novelty is that it is now possible to obtain from ITALIC lists of species potentially occurring under specified conditions, which are used for "feeding" FRIDA to generate personalized, user-oriented interactive identification key on demand.

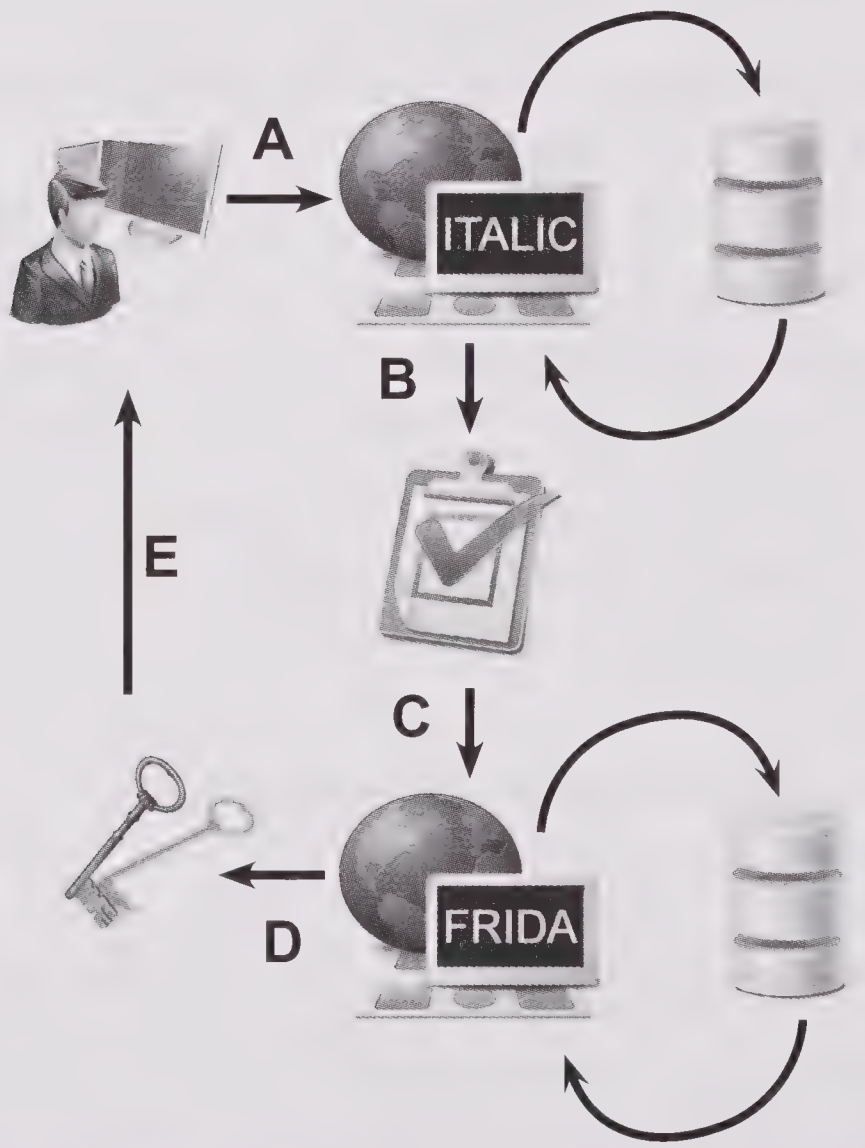


Fig. 1 – A) The user inputs in ITALIC a combination of parameters defining a "virtual habitat", B) ITALIC produces a list of species which are likely to occur under the specified conditions, C) the list is automatically sent to FRIDA, D) FRIDA produces an interactive key to the species in the list, E) the key, in its various versions (online, stand-alone for CD-ROM and for mobile devices) is forwarded to the user.

This novelty can have relevant applicative implications. For example, forest managers, which are in charge of monitoring epiphytic lichen diversity in spruce forests of the Italian administrative region Trentino-Alto Adige, could obtain their potential species list by selecting the administrative region, the 'subalpine' bioclimatic belt, 'epiphytic' lichens growing on 'acid bark' (pH indicator value

1-2), in 'mesic' (air humidity indicator value 2-4), 'shaded' (indicator value for solar irradiation 2-3), and 'non-eutrophicated' (indicator value for eutrophication 1-2) conditions. An interactive identification key to these species, including high-resolution pictures and ecological notes, is immediately created by FRIDA, and can be used for routine activities.

Similarly, managers in charge of monitoring the effectiveness of a restoration program on a Greek temple near the town of Agrigento (Sicily, S Italy), could obtain the list and the identification key for the species potentially occurring in that environment. They should just select 'Sicily', 'dry-mediterranean' bioclimatic region, 'saxicolous' lichens growing on limestone ('basic' substrata, indicator value for pH 5), in 'dry' (air humidity indicator 4-5), and 'sun-exposed' (indicator value for solar irradiation 4-5) conditions.

The digital identification keys are immediately ready to be used online. However, when identification has to be carried out in the field, the Web could be not the best medium for a key. For this reason, several accessory tools were developed, which permit the storage of the keys as stand-alone packages on different media, such as mobile devices (iPhones and other smartphones), as well as on paper, in the form of printable, illustrated field guides.

6 Conclusion

The identification of lichens is often difficult, both in the field and in the laboratory, and requires a long period of study and training. The new digital keys "on demand", being restricted to a relatively small number of species potentially occurring in a given habitat and/or in a given area, are generally much more user-friendly. They proved to be a valuable support for the technical personnel of Environmental Agencies, Nature Parks, Cultural Heritage Conservation Agencies, etc. Within *KeyToNature* we have produced free of charge hundreds of mini-keys for schools using the same approach. However, the potential market for such a service in the field of practical applications is quite high: we have already received dozen of requests by institutions and companies which are ready to pay for having the possibily of generating "their own" keys on demand.

Acknowledgement

This paper was produced in the framework of the project *KeyToNature*, funded under the *e*Content*plus* programme, a multi-annual Community programme to make digital content in Europe more accessible, usable and exploitable. — Contract no. ECP-2006-EDU-410019.

References

[1] P. L. Nimis, C. Scheidegger and P. A. Wolseley, "An introduction". In: P. L. Nimis, C. Scheidegger, and P. A. Wolseley (eds.), *Monitoring with Lichens - Monitoring Lichens.* NATO Science Series. IV. Earth and Environmental Sciences, 7, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[2] G. Caneva, M. P. Nugari and O. Salvadori, *Plant biology for Cultural Heritage*. Getty Conservation Institute, Los Angeles, 400 pp., 2008.

[3] M. R. D. Seaward, C. Giacobini, M. R. Giuliani and A. Roccardi, "The role of lichens in the

biodeterioration of ancient monuments with particular reference to Central Italy". *International Biodeterioration and Biodegradation*, vol. 48, pp. 202-208, 2001.

[4] J. Nascimbene, O. Salvadori and P. L. Nimis, "Monitoring lichen recolonization on a restored calcareous statue". *Science of Total Environment*, vol. 407, pp. 2420-2426, 2009.

[5] P. L. Nimis, *The lichens of Italy. An annotated catalogue*. Museo Regionale Scienze Naturali, Torino, Monografie, XII, 897 pp., 1993.

[6] P. L. Nimis and S. Martellos, *ITALIC - The Information System on Italian Lichens.* Version 4.0. University of Trieste, Department of Biology, IN4.0/1 (http:// dbiodbs.univ.trieste.it/), 2010.

[7] IUCN, *IUCNRed List Categories and Criteria: Version 3.1*. IUCN Species Survival Commission. Gland, Cambridge: IUCN. 30 pp., 2010.

[8] P. L. Nimis and S. Martellos, "Testing the predictivity of ecological indicator values. A comparison of real and "virtual" relevés of lichen vegetation". *Plant Ecology*, vol. 157, pp. 165-172, 2001.

[9] P. L. Nimis, Department of Life Science, University of Trieste, 2010. (personal communication)

[10] J. Nascimbene, S. Martellos and P. L. Nimis, "Epiphytic lichens of tree-line forests in the Central-Eastern Italian Alps and their importance for conservation". *The Lichenologist*, vol. 38, pp. 373-382, 2006.

[11] J. Nascimbene, L. Marini, R. Motta and P. L. Nimis, P. L., "Influence of tree age, tree size nd crown structure on lichen communities in mature Alpine spruce forests". *Biodiversity Conservation,* vol. 18, pp. 1519–1522, 2009.

[12] S. Martellos, "Multi-authored interactive identification keys: The FRIDA (FRiendly IDentificAtion) package". *Taxon,* vol. 59(3), pp. 922-929, 2010.

[13] L. Poldini, *Nuovo Atlante corologico delle piante vascolari nel Friuli-Venezia Giulia*. Reg. Auton. Friuli Ven. Giulia, Udine, 529 pp., 2010.

[14] F. Festi and F. Prosser, *La Flora del Parco Naturale Paneveggio Pale di San Martino. Atlante corologico e repertorio delle segnalazioni.* Supplemento agli Annali del Museo Civico di Rovereto, vol. 13, 438 pp., 2000.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 157-162.
ISBN 978-88-8303-295-0. EUT, 2010.

"Flora Italiana Digitale": an interactive identification tool for the Flora of Italy

Riccardo Guarino, Sabina Addamiano, Marco La Rosa, Sandro Pignatti

Abstract — The digital facilities of the second edition of Pignatti's "Flora d'Italia" are presented. A software, called FID (i.e. "Flora Italiana Digitale") will link together a random-access interactive identification tool, a thesaurus, synoptic tables and one template for each single species, including a distribution map (referred to the Italian regions), "ecograms", a text-box and up to 24 high-resolution colour images. The FID follows a "shareware phylosophy". All contents and images can be integrated and/or replaced over time, in order to continuously improve the diagnostic and qualitative performance of the provided utilities. Ideally, the community of users should interact on the web, so that every user could easily become content provider.

Index Terms — Flora of Italy, interactive identification tool.

1 Introduction

The most recent flora of the vascular plants growing in Italy was published in 1982 [1]. This work consists of 2324 pages in three volumes, where each of 5599 native and invasive species is briefly described and illustrated by a regional distribution map and a drawing, the latter mostly taken from the "Iconographia Florae Italicae" [2].

Since 1982, several changes occurred both in the systematics of vascular plants and in the floristic exploration of the Country: the records for the Italian flora raised up to 6700 species [3] and the bulk of knowledge on a single species includes on one hand more detailed information on ecological and phytosociological preferences, and on the other hand on molecular and phylogenetic data and results. Moreover, the public interest and concern for nature and the biosphere, of which vascular plants are the most visible and perceivable component (at least for most of the terrestrial ecosystems), has

R. Guarino is with the Dept. of Botanical Sciences, University of Palermo, I-90123. E-mail: riccardo.guarino@unipa.it.
S. Addamiano, viale P. Pellini 31, Perugia, I-06124. E-mail: sabinaad@tin.it.
M. LaRosa, via P. Maioli, 36, San Miniato (PI), I-56028. E-mail: marcolarosa@marcolarosa.it.
S. Pignatti is with the Dept. of Plant Biology, University of Rome "La Sapienza", I-00165. E-mail: sandro.pignatti@gmail.com.

increased consistently in the last three decades also among non-specialists [4].

For these reasons, a second edition of the Pignatti's "Flora d'Italia" was planned, making use of the new facilities offered by information technology, in order to provide an updated inventory where specialists and non-specialists can easily find the information they search for.

The new work will consist of four volumes with integrated digital utilities and data-sources, that link together interactive polytomous keys, a thesaurus, synoptic tables and one template for each single species, including a distribution map (presence-absence in the Italian regions), "ecograms", a text-box and up to 24 high-resolution colour images.

2 Digital Utilities for the Italian Flora

Our project started in December 2002. A software, called FID (i.e. "Flora Italiana Digitale"), has been written in Visual Basic and developed with the inclusion of client packages for multiplatforms.

The software aims at:

- providing a rich iconography for each of the species of the Italian Flora. Current settings are limited to max. 24 images/species, but there is virtually no limit to this number. Images may include digital photographs, optical and microscopic scannings of diagnostic characters, drawings and any other depiction that turns out to be useful to identify a given species and to represent its morphologic variability;
- helping beginners to identify the observed specimens by means of a random-access interactive identification tool, filtering the species on the basis of the morpho-anatomical characters selected by the user's queries;
- letting the user to personalize the contents of the FID, by adding/replacing texts and/or images, as well as personal data-bases (e.g. own herbarium, the flora observed during a hike...). These additional contents can be kept "private only" or, instead, submitted to the scientific committee to be placed at any other user's disposal.
- providing exhaustive open-ended descriptions on the species belonging to the Italian vascular flora, including phylogenetic, morpho-anatomical, ecological data, conservation issues, as well as information on traditional uses, common and vernacular names, adversities, importance for human economy, etc.

Being the digital flora an open-ended work, all contents and images can be integrated and/or replaced over time, in order to continuously improve the diagnostic and qualitative performance of the provided utilities. Ideally, the community of users should interact via web, so that every user could easily become content-provider, by giving the right to use the provided content in the frame of the FID. A scientific committee will evaluate the contributions prior to publication and ensure the visibility of each single contributor. Official updates will be periodically published online.

Up to now, the digital utilities of the second edition of the "Flora d'Italia" are the result of the liberal cooperation of more than 150 contributors. One relevant point is the direct involvement of secondary schools for testing polytomous keys and

usability of the contents. A second relevant point is the lack of sponsors, so that all contents will not only make more accessible the information on plant species, but also celebrate the praiseworthy synergy of people sharing the same passion for the beauty of floristic research. In this "shareware phylosophy", a mutual aid that has to be particularly mentioned is with the *Dryades* Project, the Italian branch of the European project *KeyToNature* [5]: several contributors collaborate to both projects and the exchange of know-how and visual contents greatly helped in the development of the FID.

3 ...How do they work?

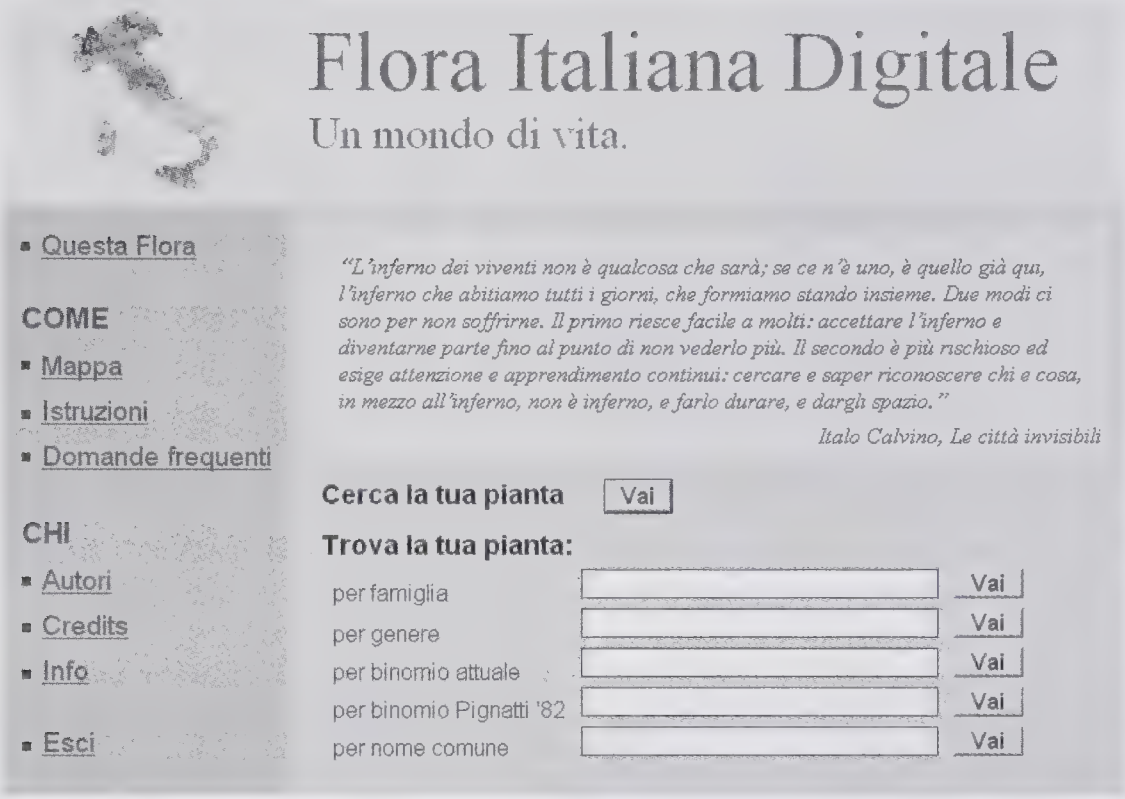

Flora Italiana Digitale
Un mondo di vita.
Questa Flora
COME
Mappa
Istruzioni
Domande frequenti
CHI
Autori
Credits
Info
Esci
"L'inferno dei viventi non è qualcosa che sarà; se ce n'è uno, è quello già qui, l'inferno che abitiamo tutti i giorni, che formiamo stando insieme. Due modi ci sono per non soffrirne. Il primo riesce facile a molti: accettare l'inferno e diventarne parte fino al punto di non vederlo più. Il secondo è più rischioso ed esige attenzione e apprendimento continui: cercare e saper riconoscere chi e cosa, in mezzo all'inferno, non è inferno, e farlo durare, e dargli spazio."
Italo Calvino, Le città invisibili
Cerca la tua pianta
Vai
Trova la tua pianta:
per famiglia
per genere
per binomio attuale
per binomio Pignatti '82
per nome comune
Vai

Fig. 1 – Opening window of the FID: the options under "trova la tua pianta" (= find your plant) can be used to go directly to the information on a plant known by the user. Instead, the option "Cerca la tua pianta" (= search your plant) opens the window of the interactive identification tool.

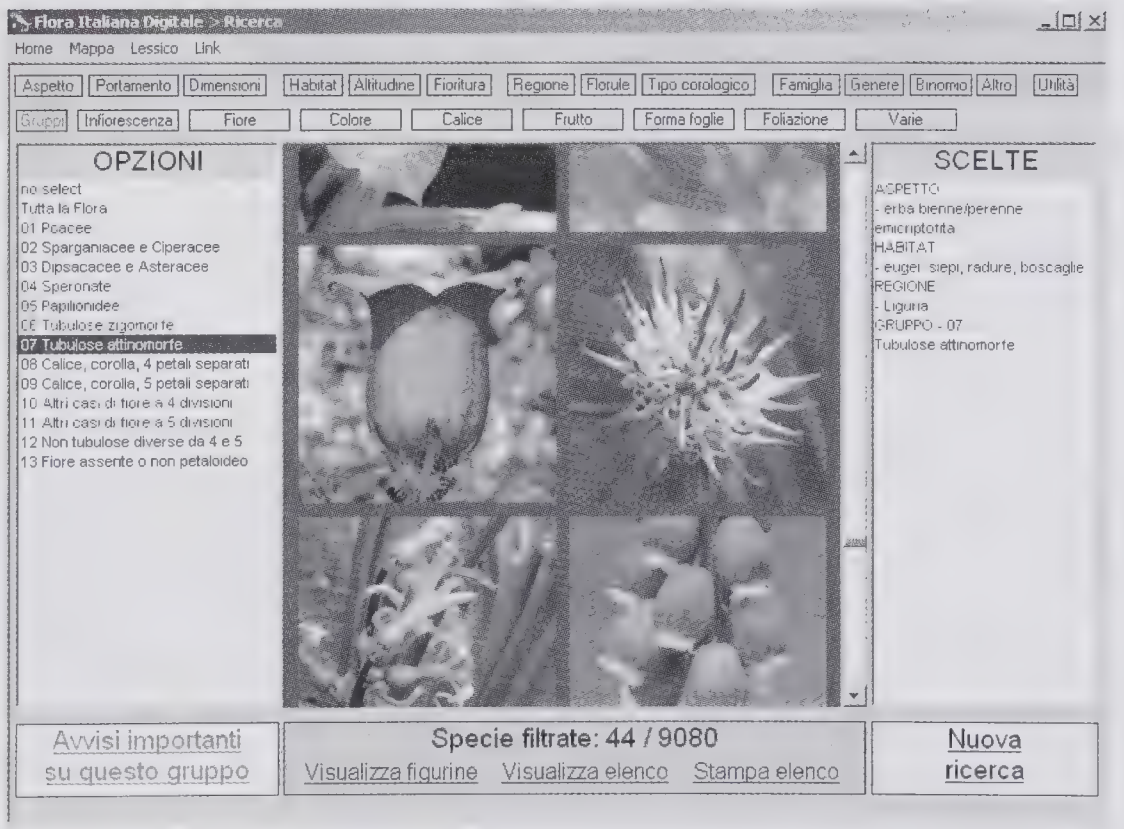

Flora Italiana Digitale > Ricerca
Home Mappa Lessico Link
Aspetto Portamento Dimensioni Habitat Altitudine Fioritura Regione Flonule Tipo corologico Famiglia Genere Binomio Altro Utilità
Gruppi Infiorescenza Fiore Colore Calice Frutto Forma foglie Foliazione Varie
OPZIONI
no select
Tutta la Flora
01 Poacee
02 Sparganiacee e Ciperacee
03 Dipsacacee e Asteracee
04 Speronate
05 Papilionidee
06 Tubulose zigomorfe
07 Tubulose attinomorfe
08 Calice, corolla, 4 petali separati
09 Calice, corolla, 5 petali separati
10 Altri casi di fiore a 4 divisioni
11 Altri casi di fiore a 5 divisioni
12 Non tubulose diverse da 4 e 5
13 Fiore assente o non petaloideo
SCELTE
ASPETTO
- erba bienne/perenne
emicriptofita
HABITAT
- eugeo: siepi, radure, boscaglie
REGIONE
- Liguria
GRUPPO - 07
Tubulose attinomorfe
Avvisi importanti su questo gruppo
Specie filtrate: 44 / 9080
Visualizza figurine Visualizza elenco Stampa elenco
Nuova ricerca

Fig. 2 – To identify a specimen, users can select a set of non-hierarchized fields and options. User's choices are listed in the left. The filtered species appear by clicking on "Visualizza figurine" (images) or "Visualizza elenco" (names, printable with "Stampa elenco"). The centre hosts a table with texts and images (1496 in total).

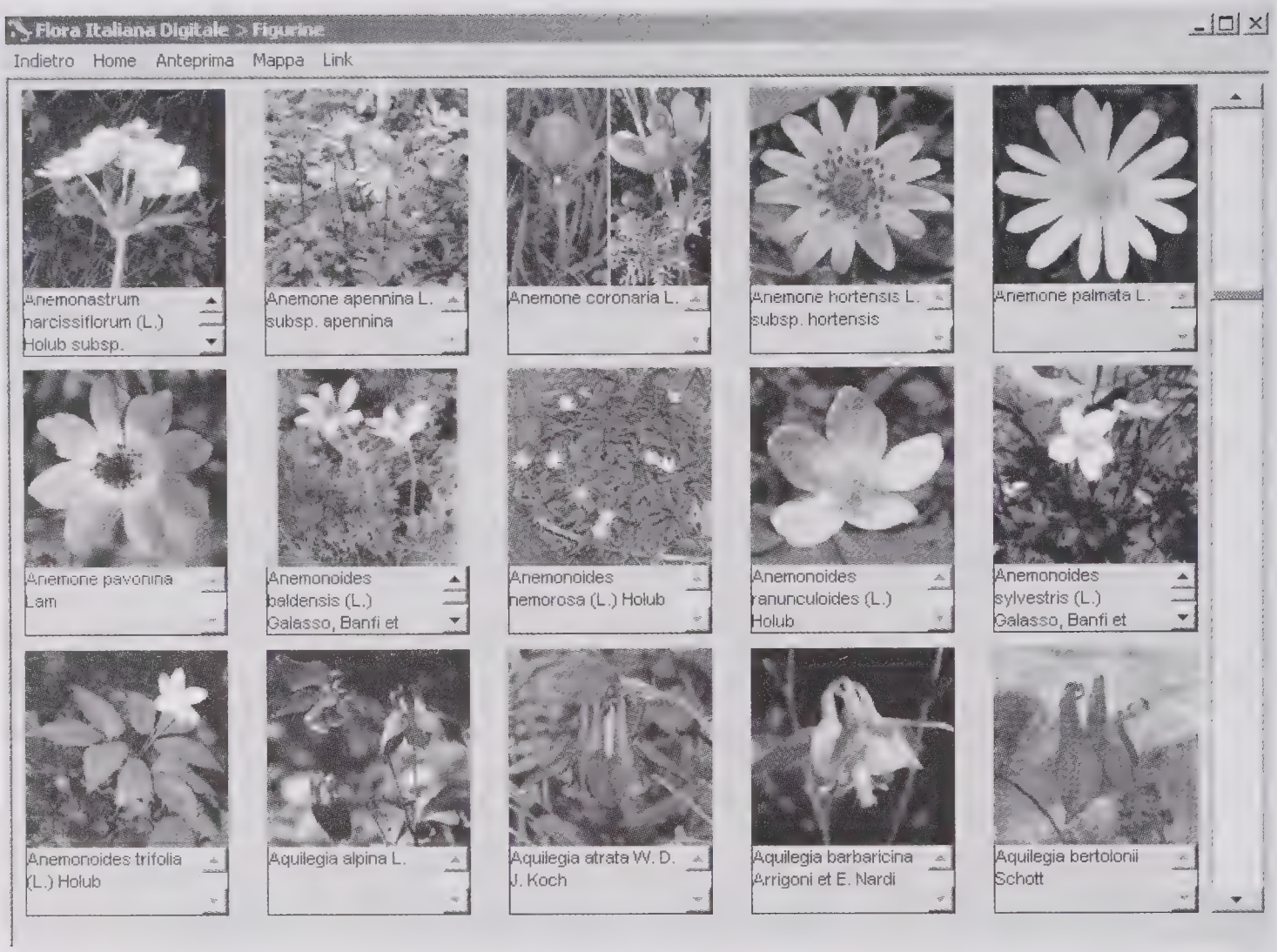

Flora Italiana Digitale > Figurine
Indietro Home Anteprima Mappa Link
Anemonastrum narcissiflorum (L.) Holub subsp.
Anemone apennina L. subsp. apennina
Anemone coronaria L.
Anemone hortensis L. subsp. hortensis
Anemone palmata L.
Anemone pavonina Lam
Anemonoides baldensis (L.) Galasso, Banfi et
Anemonoides nemorosa (L.) Holub
Anemonoides ranunculoides (L.) Holub
Anemonoides sylvestris (L.) Galasso, Banfi et
Anemonoides trifolia (L.) Holub
Aquilegia alpina L.
Aquilegia atrata W. D. J. Koch
Aquilegia barbaricina Arrigoni et E. Nardi
Aquilegia bertolonii Schott

Fig. 3 – The visual identification of a specimen is possible through the comparison of thumbnails. A single click will magnify the image. A double-click will open the informative window of each single species.

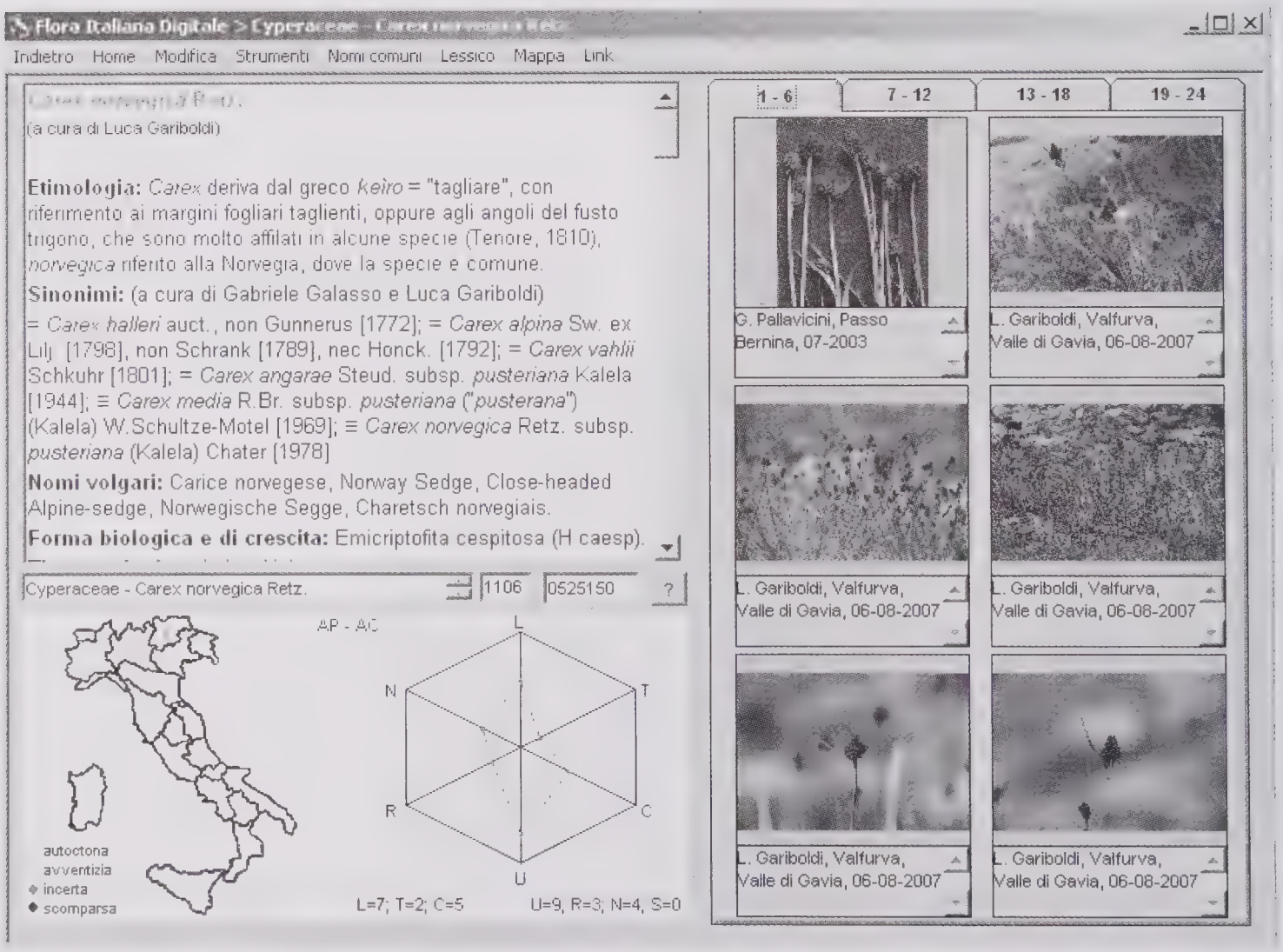

Flora Italiana Digitale > Cyperaceae
Indietro Home Modifica Strumenti Nomi comuni Lessico Mappa Link
(a cura di Luca Gariboldi)
Etimologia: Carex deriva dal greco keiro = "tagliare", con riferimento ai margini fogliari taglienti, oppure agli angoli del fusto trigono, che sono molto affilati in alcune specie (Tenore, 1810); norvegica riferito alla Norvegia, dove la specie è comune.
Sinonimi: (a cura di Gabriele Galasso e Luca Gariboldi)
= Carex halleri auct., non Gunnerus [1772]; = Carex alpina Sw. ex Lilj [1798], non Schrank [1789], nec Honck. [1792]; = Carex vahlii Schkuhr [1801]; = Carex angarae Steud. subsp. pusteriana Kalela [1944]; ≡ Carex media R.Br. subsp. pusteriana ("pusterana") (Kalela) W.Schultze-Motel [1969]; ≡ Carex norvegica Retz. subsp. pusteriana (Kalela) Chater [1978]
Nomi volgari: Carice norvegese, Norway Sedge, Close-headed Alpine-sedge, Norwegische Segge, Charetsch norvegiais.
Forma biologica e di crescita: Emicriptofita cespitosa (H caesp).
Cyperaceae - Carex norvegica Retz.
1106
0525150
autoctona
avventizia
incerta
scomparsa
AP - AC
L
N
T
R
C
U
L=7; T=2; C=5
U=9, R=3; N=4, S=0
1 - 6
7 - 12
13 - 18
19 - 24
G. Pallavicini, Passo Bernina, 07-2003
L. Gariboldi, Valfurva, Valle di Gavia, 06-08-2007
L. Gariboldi, Valfurva, Valle di Gavia, 06-08-2007
L. Gariboldi, Valfurva, Valle di Gavia, 06-08-2007
L. Gariboldi, Valfurva, Valle di Gavia, 06-08-2007
L. Gariboldi, Valfurva, Valle di Gavia, 06-08-2007
L. Gariboldi, Valfurva, Valle di Gavia, 06-08-2007
L. Gariboldi, Valfurva, Valle di Gavia, 06-08-2007

Fig. 4 – The informative window includes: a standard text (that can be modified/ personalized by the user); a distribution map; an "ecogram" [6] displaying the ecological preferences and pollination/dissemination strategies; the thumbs of up to 24 high resolution images.

Ranunculaceae - Anemone palmata L
R. Guarino, Orto Botanico, Cagliari, 04-02-2005, culta
Ranunculaceae - Anemone palmata L
G. Campus, Corongiu, Cagliari, 03-03-2008
Ranunculaceae - Anemone palmata L
G. Campus, Corongiu, Cagliari, 13-03-2008
Ranunculaceae - Ranunculus ficaria L
M. La Rosa, Fontevivo, San Miniato, 31-03-2005
Ranunculaceae - Ranunculus ficaria L
M. La Rosa, Fontevivo, San Miniato, 31-03-2005
Ranunculaceae - Ranunculus ficaria L
V. Tomaselli, Dolomiti Lucane, Potenza, 02-05-2005
Posizione 4
Modifica
Zoom
Start
MISURE DI TUTELA E CO...
Flora Italiana Digitale > ...
Tavolozza

Fig. 5 – It is possible to collect up to six selected images in a synoptic table, in order to compare different features of one or more species. Each image can be zoomed; colours, contrast and light can be temporarily modified by the user, in order to better observe diagnostic characters.

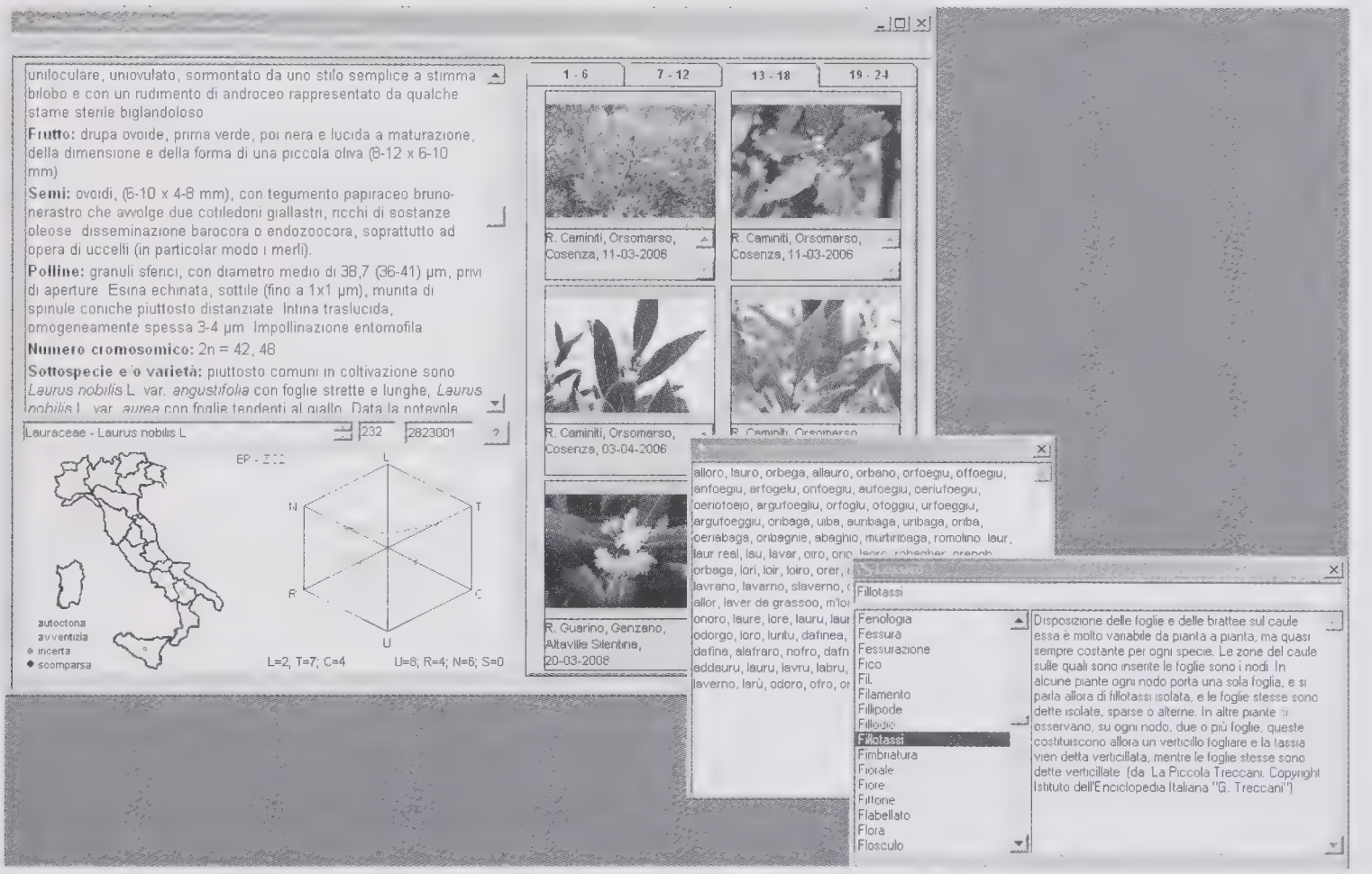

unilocolare, uniovulato, sormontato da uno stilo semplice a stimma bilobo e con un rudimento di androceo rappresentato da qualche stame sterile biglandoloso
Frutto: drupa ovoide, prima verde, poi nera e lucida a maturazione, della dimensione e della forma di una piccola oliva (8-12 x 6-10 mm)
Semi: ovoidi, (5-10 x 4-8 mm), con tegumento papiraceo bruno-nerastro che avvolge due cotiledoni giallastri, ricchi di sostanze oleose disseminazione barocora o endozoocora, soprattutto ad opera di uccelli (in particolar modo i merli).
Polline: granuli sferici, con diametro medio di 38,7 (36-41) µm, privi di aperture Esina echinata, sottile (fino a 1x1 µm), munita di spinule coniche piuttosto distanziate Intina traslucida, omogeneamente spessa 3-4 µm Impollinazione entomofila
Numero cromosomico: 2n = 42, 48
Sottospecie e/o varietà: piuttosto comuni in coltivazione sono Laurus nobilis L. var. angustifolia con foglie strette e lunghe, Laurus nobilis L. var. aurea con foglie tendenti al giallo Data la notevole
Lauraceae - Laurus nobilis L
232
2823001
1 - 6
7 - 12
13 - 18
19 - 24
R. Caminiti, Orsomarso, Cosenza, 11-03-2006
R. Caminiti, Orsomarso, Cosenza, 11-03-2006
R. Caminiti, Orsomarso, Cosenza, 03-04-2006
R. Guarino, Genzano, Altavilla Silentina, 20-03-2008
autoctona
avventizia
incerta
scomparsa
L=2, T=7; C=4
U=8; R=4; N=6; S=0
alloro, lauro, orbega, allauro, orbano, orfoegiu, offoegiu, anfoegiu, arfogelu, onfoegiu, aufoegiu, oeriufoegiu, oeriofoeio, argufoegiu, orfoglu, ofoggiu, urfoeggiu, argufoeggiu, oribaga, uiba, auribaga, uribaga, oriba, oeriabaga, oribagnie, abaghio, murtiribaga, romolino laur, laur real, lau, lavar, oiro, orio, laors, rabagher, orenab
lavrano, lavarno, slaverno,
allor, laver de grassoo, m'lo
onoro, laure, lore, lauru, laur
odorgo, loro, luritu, dafinea,
dafina, alafraro, nofro, dafn
addauru, lauru, lavru, labru,
laverno, larù, odoro, ofro, or
Fillotassi
Fenologia
Fessura
Fessurazione
Fico
Fil.
Filamento
Fillipode
Fillodio
Fillotassi
Fimbriatura
Fiorale
Fiore
Fittone
Flabellato
Flora
Flosculo
Disposizione delle foglie e delle brattee sul caule essa è molto variabile da pianta a pianta, ma quasi sempre costante per ogni specie. Le zone del caule sulle quali sono inserite le foglie sono i nodi In alcune piante ogni nodo porta una sola foglia, e si parla allora di fillotassi isolata, e le foglie stesse sono dette isolate, sparse o alterne. In altre piante si osservano, su ogni nodo, due o più foglie, queste costituiscono allora un verticillo fogliare e la tassia vien detta verticillata, mentre le foglie stesse sono dette verticillate (da La Piccola Treccani. Copyright Istituto dell'Enciclopedia Italiana "G. Treccani")

Fig. 6 – Some facilities, such as a conceptual map, a thesaurus, the list of common names can be recalled by the user at any time, in order to make browsing more friendly.

Acknowledgement

The authors gratefully acknowledge all the researchers and colleagues that liberally provided information and images to the Digital Flora of Italy. Without their passionate support it would not have been possible to collect the (up to now) 80000 digital images included in our work.

References

[1] S. Pignatti, *Flora d'Italia.* Edagricole, Bologna, vol. 1-3, 1982.

[2] A. Fiori, *Iconographia Florae Italicae*, *ossia Flora Italiana Illustrata*. 3rd ed. Tipografia Editrice Mariano Ricci, Firenze, 1933.

[3] F. Conti, G. Abbate, A. Alessandrini and C. Blasi (eds.), *An annotated Checklist of the Italian Vascular Flora*. Palombi Editori, Roma, 2005.

[4] R. Guarino, S. Addamiano, M. La Rosa and S. Pignatti. "The impact of Information Technology on the identification of species and archiviation of taxonomic and floristic data". *Bocconea,* vol. 23, pp.19-23, 2009.

[5] P. L. Nimis and S. Martellos, "KeyToNature – Dryades" http://www.dryades.eu/home1.html, 2008.

[6] S. Pignatti, H. Ellenberg and S. Pietrosanti, "Ecograms for phytosociological tables based on Ellenberg's Zeigerwerte". *Annali di Botanica,* vol. 54, pp. 5-14, 1996.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 163-169.
ISBN 978-88-8303-295-0. EUT, 2010.

eFlora and DialGraph, tools for enhancing identification processes in plants

Fernando Sánchez Laulhé, Cecilio Cano Calonge,
Antonio Jiménez Montaño

Abstract—This document describes in some detail the eFlora web application, a powerful tool for the identification of plant species. It incorporates the corpus of Flora Iberica, a scientific description of the vascular plants living in the Iberian Peninsula, which is treated as unstructured information and therefore indexed by a full text search engine tool, in our case Lucene. eFlora also includes dichotomous keys, which are displayed using Hyperbolic Geometry. By making intelligent use of the keys, we have created two original and useful features, the comparison of arbitrarily chosen species, which is resolved by a dynamic generation of subkeys applied to these selected species, and the presentation of dichotomous keys in the form of a Virtual Assistant, or conversational robot, using our solution DialGraph, which allows to non-academic users an approach in Natural Language, such as chat, or voice recognition, Text to Speech Synthesis (TTS) or even Automatic Translation when dealing with a multilanguage context. Concerning the configuration of the Virtual Assistant, we provide a very intuitive BPM-like graphical design. This approach to dichotomous keys helps teaching biodiversity science, enhances the awareness of its importance, and makes citizens emotionally closer to science.

Index Terms — virtual assistant, hyperbolic geometry, natural language, dichotomous key, Bot, automatic translation, Flora Iberica, BPM, Text to Speech Synthesis, TTS.

1 Introduction

Nowadays there is a major challenge to properly process the large body of available botanical information, most of which lacks structure consensus, probably due to both its ancient origins and the inherent difficulty to structure it in a consistent manner. This document describes in some detail the eFlora web application, a powerful tool for the identification of plant species, which incorporates the corpus of Flora Iberica, a scientific description of the vascular plants living in the Iberian Peninsula. eFlora has been addressed

The authors are with Terasoft Consultores S.L., Parque Empresarial Punta Galea c/.Perú, 4. Loft 3 28230 Las Rozas (Madrid), Spain. E-mail of first author: fsl@terasoft.es.

with a full text type approach, and is accessible via an appropriate open source search engine, Lucene, which allows to launch free text queries, the system responding with a list of species, ranked in order of relevance. Being Flora Iberica a work aimed at the identification of species, we have also included dichotomous keys in graphic form and under hyperbolic geometry. Furthermore, the original functionality of these keys has been enhanced in the form of dynamic generation of sub-keys from any arbitrary selection of species. Another intesting feature, implemented with a dichotomous key to Iberian Conifers, is the presentation of the key in the form of a conversational robot or virtual assistant, that interacts with the user in natural language.

2 Flora Iberica

2.1 Descriptions

To give some structure to the information contained in Flora Iberica, while maintaining the characteristic of free natural language descriptions, the description part of each species has been separated into a priori-defined fields (Tab. 1).

Level 1 fields	*Level 0 fields*
Plant	Root
	Basal stem
	Stems and branches
	Spines
Leaves	Leaves
Flower	Inflorescence
	Receptacle
	Calyx
	Corolla
Gynaeceum	Gynaeceum
Androecium	Androecium
	Nectary
Fruit	Fruit
	Seed

Tab. 1 – Fields used for organising the descriptions of species in Flora Iberica.

These fields can be grouped into even greater granularity, in anticipation that some of them may belong to categories of higher order. For example, a query can be designed as generically directed to the field *Flower*, or it may be more detailed, specifying which field in the category *Flower* the query is to be launched: *Inflorescence*, *Receptacle*, *Calyx* and *Corolla*. In this way, queries can be made either under both full free text scheme, or directed, in free text as well, towards specific fields.

2.2 Dichotomous keys by hyperbolic geometry

The dichotomous keys of Flora Iberica have been digitized. These keys are specific for all taxonomic levels used in the original work i.e. there are special keys that identify species within the same genus, keys to identify genera within a family, etc.

The dichotomous keys have been ported to XML format for processing in the form of trees in hyperbolic geometry, whose root concept is that it allows to pass through a point more than one line parallel to another line that does not pass through that point. With respect to information representation - decision trees in our case – this means that the hyperbolic space can be represented as a circle in which the periphery represents infinity. In this geometry, the closer to the edge (infinity) we are, the smaller is the size of what we represent. This allows to represent graphs in hyperbolic geometry maintaining the properties of focus and context. Everything in the center is large, and if we move aside to the periphery, it becomes smaller, but still visible. For the implementation of hyperbolic trees, Treebolic, an open source solution, has been used (Fig. 1).

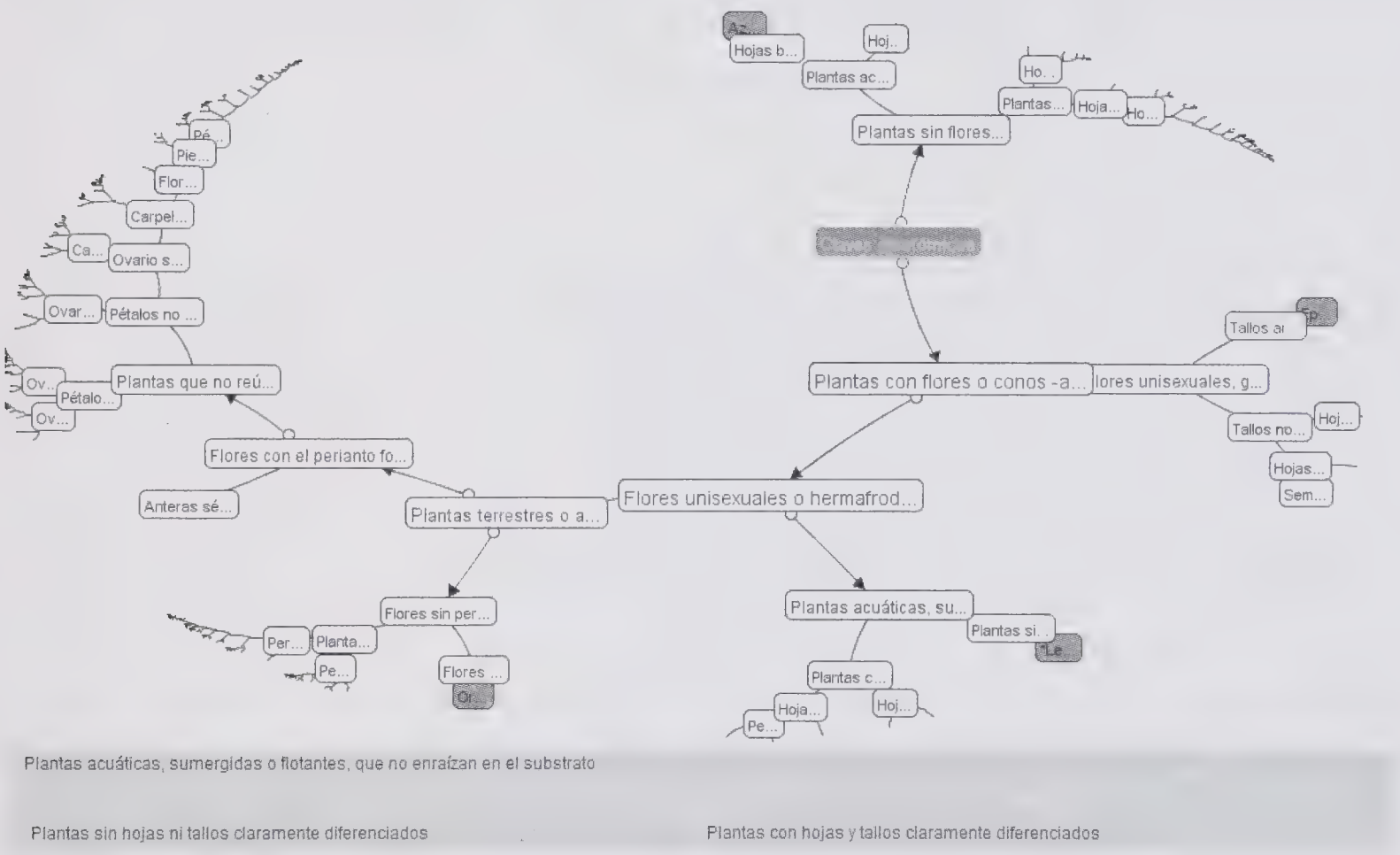


Fig. 1 – General view of dichotomous keys at initial execution from the root node, displayed using hyperbolic geometry.

3 Indexing

Lucene, an open source solution from Apache Group, was used as the main search engine. We have created a data model that includes a set of indexed fields specifically for using in queries. Other fields, such as Flowering periods, Observations or Bibliography are used only as supplementary information to be shown when displaying a full description.

As for the ability to query in English, we have added a filter that acts between

the user input string and the search engine. In this way, a user not speaking Spanish, but able to understand botanical information, can query directly in English.

4 Identification

4.1 By species comparison

While eFlora indexes the data with a full text search engine, queries have a qualitative difference in relation to those made in a classic form, such as in Google. In the latter case, some information is sought, and the result can be found in one or more points from the list of results provided by Google. It is the surfer who decides what information is to be considered as valid. This means that sometimes useful information may be scattered among a high number of returned results. In eFlora the situation is completely different. The question is: what is the species that we have in our hands? For anwering, it is not sufficient for the search system to just offer a series of results ranked by relevance. This is useful, but what is necessary is that the system can tell us what to observe for differentiating between the items included in the list of results. Otherwise the user should make a very tedious work examining the full description of every species in the list, trying to find subtle differences among similar taxa.

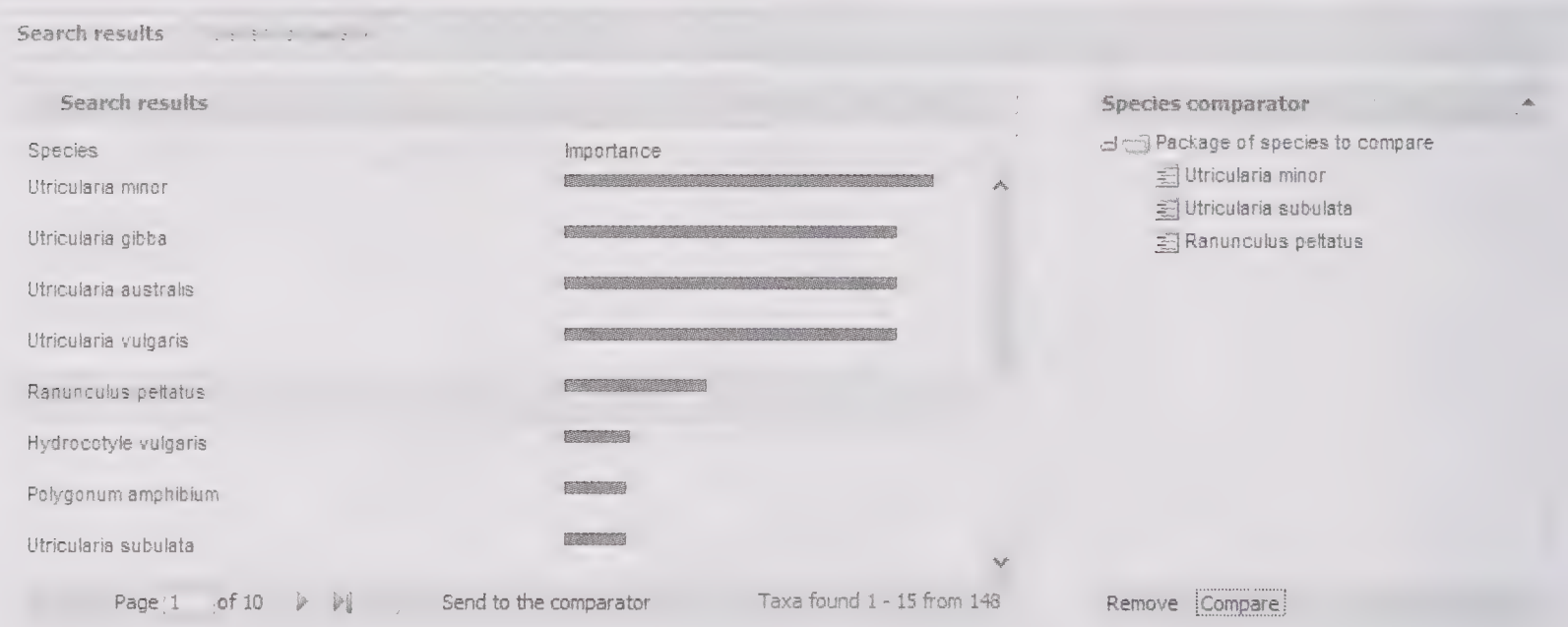


Fig. 2 – Example with a list of results ranked by relevance; on the right side, the repository filled up with three species to be compared.

With eFlora the identification task, which is originated through a query using terms describing certain observable characteristics of a certain species, begins with a set of species that have a good chance of containing the targeted specimen. To differentiate among taxa, users can select those which look more likely, and bring them to a special repository in which they can perform the function of comparison (Fig. 2). This function processes the entire tree of the dichotomous key in real time and extracts a subkey that shows the differences among the selected species. In this way, the user has a useful tool to identify the species in question (Fig. 3).

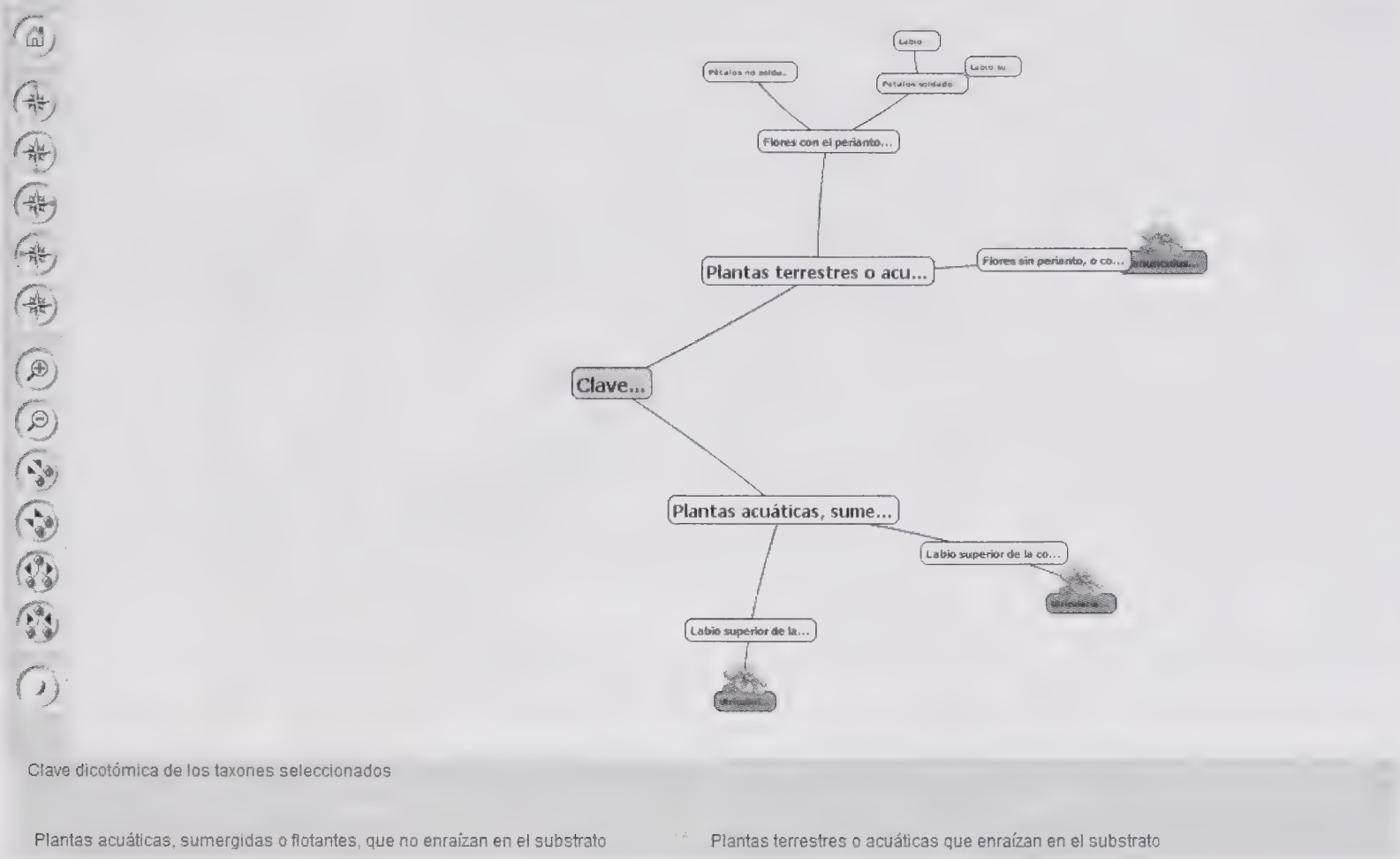

Labio
Flores con el perianto...
Plantas terrestres o acu...
Flores sin perianto, o co...
Clave...
Plantas acuáticas, sume...
Labio superior de la co...
Labio superior de la...
Clave dicotómica de los taxones seleccionados
Plantas acuáticas, sumergidas o flotantes, que no enraízan en el substrato
Plantas terrestres o acuáticas que enraízan en el substrato

Fig. 3 – Once the compare function has been executed, a dynamic subkey is displayed in hyperbolic geometry.

4.2 The Virtual Assistant

We have developed an original tool for teaching the identification of biodiversity, based on the use of Virtual Assistants in the form of a conversational bot. A Virtual Assistant is a software application capable of maintaining a conversation with users on a specific topic in natural language. We applied DialGraph to a dichotomous key of Iberian Conifers. DialGraph allows an almost complete portability between dichotomous keys - considered as a large tree structure - and its graphical system design. It allows communication with users through Chat, materialized in a virtual character who pronounces sentences and might even recognize the voice of the Internet surfer. This system is topologically equivalent to the traditional dichotomous keys, but for its ability to chain multiple words appearing in the same sentence, and the freedom to use popular expressions and to apply automatic translation, it has important advantages for teaching Biodiversity, mainly when focused towards young users (Fig. 4).

The strongest point of DialGraph is the tool created for designing the conversation logic, which we call The Designer. This can be accessed through the web using standard navigators, such us IE, FF, Chrome, etc., and models the dialog by drag and drop objects within a virtual dashboard. These objects represent different assistant states and the links among states. By defining in links as parameters very simple linguistic lexical roots and appropriate messages in states, the dialog with users remains fully determined as long as the user is informed about the aim of the assistant (Fig. 5).

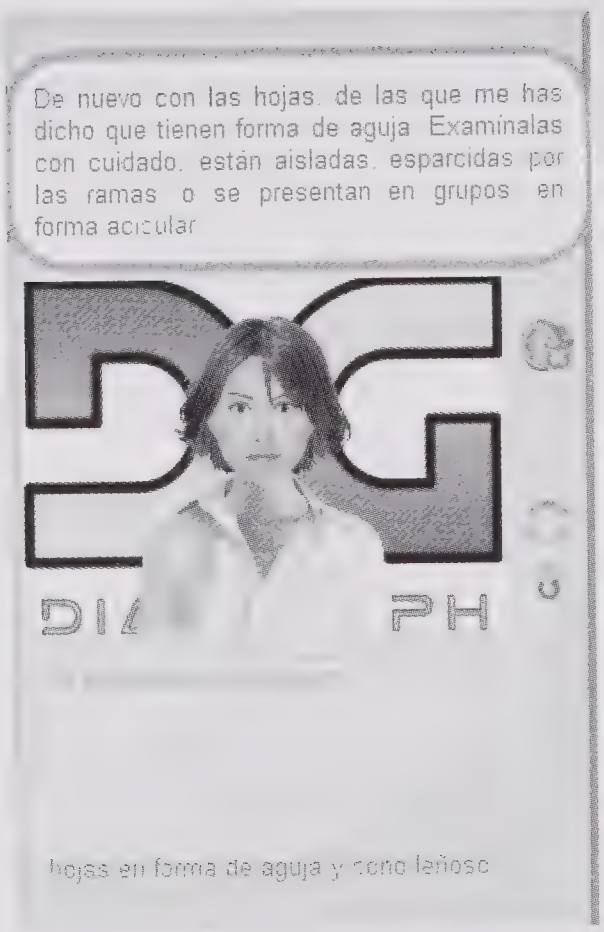

De nuevo con las hojas. de las que me has dicho que tienen forma de aguja Examínalas con cuidado. están aisladas. esparcidas por las ramas o se presentan en grupos en forma acicular
hojas en forma de aguja y cono leñoso

Fig. 4 – Moving avatar interacting with users in the web.

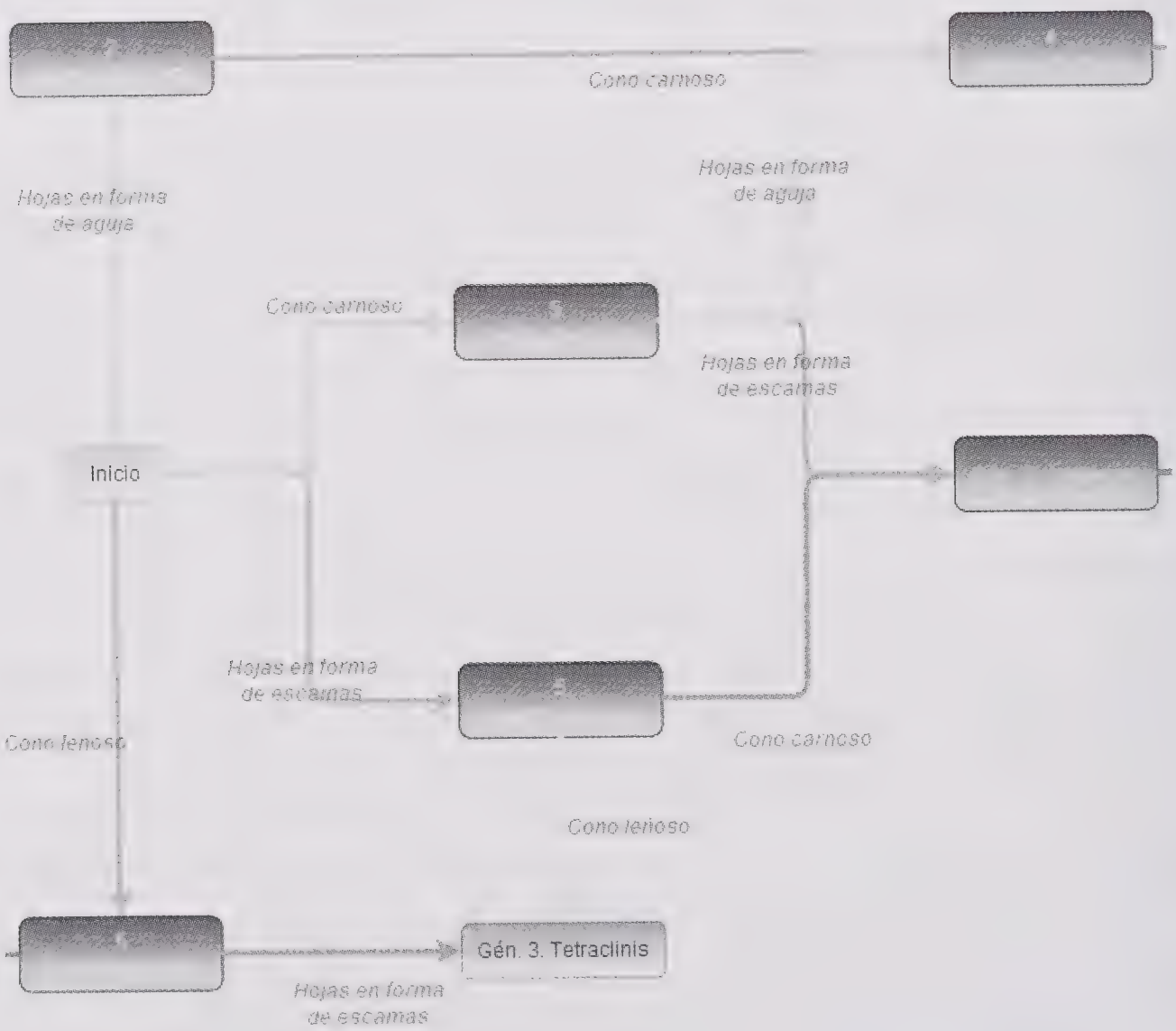


Fig. 5 – Designing the flow of dialog for a specific part of the dichotomous key using DialGraph as a BPM tool.

5 Conclusion

TeraSoft Consultores SL has created, together with the Botanic Garden of Madrid, a set of tools which enhances the identification of plant species. Based on the digitalization of botanic data “as they are” in the classic form, including dichotomous keys, it permits to approach science to citisens and to effectively teach the importance of Biodiversity.

References

[1] E-Flora Iberica: http://www.efloraiberica.es/eflora/, 2010.
[2] DialGraph: http://www.dialgraph.com, 2010.
[3] TeraSoft Consultores SL: http://www.terasoft.es, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 171-175.
ISBN 978-88-8303-295-0. EUT, 2010.

A catalogue of bird bones: an exercise in semantic web practice

Gudmundur Gudmundsson, Seth D. Brewington,
Thomas H. McGovern, Aevar Petersen

Abstract — The vast databases of natural history collections are increasingly being made accessible through the internet. The challenge is to place this data in a wider context that may reach beyond the interests of scholars only. The North Atlantic Biocultural Organization and Icelandic Institute of Natural History are jointly developing a web based catalogue of bird bones, comprising digital images, and related information from the museum database. Linking the bird bone catalogue with the semantic web developed by STERNA will integrate the bird bone catalogue with diverse information on birds that is directed towards the general public.

Index Terms — bird bones, collections, natural history, semantic web.

1 Introduction

Natural history museums have a long tradition of collecting and preserving specimens primarily for scientific study and also for public education and exhibition. Museum research has over time generated vast collections of specimens and related information, which was mostly accessible only to a closed community of researchers. The main reason was that collection information was registered on paper files and index cards. Information retrieval was thus confined to trained personnel only. Present museums almost universally rely on computer databases to register and update information of their vast collections, and almost all institutions are in the process of digitizing information on paper and card files. In addition to this, diverse information that

G. Gudmundsson is with the Icelandic Institute of Natural History, Hlemmur 3, 125 Reykjavik, Iceland. E-mail: gg@ni.is.
S. Brewington is with the Graduate School and University Canter, CUNY, New York, N.Y. E-mail: SB@colostate.edu.
T. McGovern is with the Graduate School and University Canter, CUNY, New York, N.Y. E-mail: SB@colostate.edu.mail.
A. Petersen Is with the Icelandic Institute of Natural History, Hlemmur 3, 125 Reykjavik, Iceland E-mail: aevar@ni.is.

has resulted from scientific studies on specimens is often integrated to museum databases, like morphological measurements, chemical analyses and digital photos. Many collection databases have thus grown to comprise numerous data tables that share one or more common data fields. For example, a species name is commonly present in data tables that store information on collection locality, morphological measurements, chemical analysis, photos, etc. This conceptual link between data tables creates novel possibilities of filtering out information from diverse datasets, like species diversity at some locality, correlations in species distributions, etc. However, collection databases with clearly defined relations between data sets and standardized data entries can not only be utilized for research, but may also be utilized for educational purposes through the web. [1]

Effective dissemination of information entails making information readily available in such a way that users with different needs have the ability to comprehend it. Information on museum collections should therefore be provided in layers of successively greater depth and detail and in a variety of different contexts. Such a "virtual museum" may be defined as a means to establish access, context, and outreach by using information technology and to establish interactive dialog with users. [2]

2 The STERNA Project: Dissemination of Collection Information

Over the last ten years, ever more cultural and scientific resources have been digitised and made accessible on the internet. However, integrated semantic search and access to these resources that are hosted in many heterogeneous databases is still difficult to achieve. The vision of the STERNA project is to provide cultural and scientific heritage institutions the opportunity to make their digital collections accessible in a light weight fashion. This will be achieved by setting up a distributed digital library that is based on semantic web technologies and standards, such as RDF (Resource Description Framework) and SKOS (Simple Knowledge Organisation System). STERNA is especially designed for small and medium sized institutions with limited budget and technical staff. Thirteen European cultural heritage institutions, multimedia archives, technology providers and research organizations, from ten countries, are participating in the STERNA project.

A network of (semantically) related digital resources is accomplished by connecting data provided by several institutions through a reference structure, which comprises several kinds of "controlled vocabulary", from simple word lists and glossaries to taxonomies, thesauri and ontology. The architecture of STERNA is thus based on distributed data repositories, which are semantically connected into a network that can be searched from different perspectives (faceted search). The network can also be extended by adding new members as well as tools, instruments and guidelines. Individual organisations can thus connect to a wider network of content holding organisations and place their data in a wider and more general context [3].

The STERNA project is partly funded the European Commission's

eContent*plus* programme. It started in June 2008 and will end in May 2011, and the participating institutions are:

1. Salzburg Research, Austria (Coordination)
2. Archipelagos, Greece
3. DOPPS BirdLife Slovenia
4. Heritage Malta
5. 5.Hungarian Natural History Museum
6. Icelandic Institute of Natural History
7. Nat. History Museum of the Municipality of Amaroussion Greece
8. Natural History Museum of Luxembourg
9. Naturalis, Natural History Museum of the Netherlands
10. Netherlands Institute of Sound and Vision
11. Royal Museum for Central Africa, Belgium
12. Teylers Museum, Netherlands
13. Wildscreen/ARKive, UK
14. Trezorix, NL

The data sources are of different types and sizes, from natural history museums, audiovisual archives, research institutions and nature conservation agencies. The vision is to create a dispersed network of information nodes, where each is supported and sustained by a member institution. For practical reasons (limited funds and staff) the STERNA project is focused on data access on birds, although the general objective is to extend the network to serve worldwide audience with more general interest in nature and wildlife. The aims are to:

1. Offer substantial amount of data by the combined effort of several institutions.
2. Linking the data in semantic context.
3. Providing advanced site functionalities, such as faceted search.
4. Offering possibilities for users to contribute additional data.

3 The Bird Bone Catalogue

Bird bones are a constant source of interest both to nature observers and professionals. Excavation of archaeological sites often yields rich assemblages of zoological remains. Zooarchaeological studies in Iceland have indicated that early settlement relied to a greater extent on hunting, than in later periods. Prior to the eleventh century the faunal remains include a large number of fish and wild birds in addition to domesticated animals, while in later periods the presence of wild fauna drops dramatically [4], [5], [6], [7].

Accurate identification of bird bones often requires consultation of reference specimens in natural history collections. However, this usually requires visit to the museums, which can be time consuming and expensive. Internet access to photos of zoological bone specimens and associated information facilitates proper identification. If these bone data are linked to a wider source of information on birds, it may be of use to a much larger audience than a closed research community. It is on these grounds that The Icelandic Institute of Natural History and NABO, in association with STERNA, decided to develop internet access to

a catalogue of bird bones with conceptual relations to diverse bird information of cultural significance.

NABO (North Atlantic Biocultural Organization) was founded over 20 years ago in an attempt to cross-cut national and disciplinary boundaries of researchers in several fields of studies, like archaeology, biology, geology, and anthropology. NABO has worked to aid in improving basic data comparability, in assisting practical fieldwork and interdisciplinary ventures, in promoting student training, and in dissemination of knowledge to other scholars, funding agencies, and the general public. [8]

The objective of the bone catalogue is to provide basic information on the internet to aid identification of bird bones. Photos and associated information are registered and maintained in a relational database, which comprises three primary data sets:

1. Taxonomical classification of birds, including scientific and vernacular names in many languages
2. Photographs and descriptions of the major bones in 54 species that have a long tradition of being utilized by humans and are of cultural relevance.
3. Specific descriptions of individual bones of a particular species – along with a simple general directory of the major bones in birds: e.g. the skull, bones of the wings and legs, keel (sternum), pelvic girdle, furcula and coracoid.

Associated with these primary data sets are two secondary (supportive) data tables:

1. Inventory of available reference specimens of a particular species at the IINH
2. Exhaustive registry of literature references on Icelandic bird fauna.

The bone catalogue will open on the web by the end of 2010. It is not intended as a conventional identification key, as it does not provide stepwise guidance to reach a final identification. Instead, it provides two search options when looking for images of bird bones: 1) taxon name (species, genus or family) and 2) type of bone. These can be used separately or in combination. A search that is limited to, say, one type of bone and a single genus will filter and display all images of skulls of that genus. The associated (secondary) data tables provide an optional inventory of available specimens at the IINH and a fairly complete list of literature with relevance to Icelandic populations of birds.

Information on bird bones in museum collections, literature and inventory of museum specimens are not likely to interest others than archeologists and ornithologists, with focused reasearch interest on that subject. The intention with the cooperation between STERNA, NABO and IIHN is to enrich the bone catalogue by making it a part of a diverse semantic network. Images of bird bones would then be accessible in a conceptual context to bird enthusiast, ordinary nature observers, as well as outreach and educational programmes.

Acknowledgement

The authors wish to thank Kjartan Birgisson for technical computing and setting up the prototype of the bone catalogue. We are also indebted to Þorvaldur Björnsson for his effort in selecting suitable bones for photographing.

References

[1] G. F. MacDonald, G. F. and S. Alsford, "The Museum as Information Utility". *Museum Management and Curatorship*, vol. 10, pp. 305-311, 1991.

[2] W. Scweibenz, "The Virtual Museum 1: New Perspectives For Museums to Present Objects and Information Using the Internet as a Knowledge Base and Communication System", In: H. H. Zimmermann and V. Schramm (eds.), *Knowledge Management and Kommunikationssysteme, Workflow Management, Multimedia, Knowledge Transfer. Proceedings des 6. Internationalen Symposiums für Informationswissenschaft* (ISI 1998), Prag, 3. – 7. November, pp. 185-200, 1998.

[3] A. Mulrenin, *STERNA: Semantic Web-based Thematic European Reference Network Application, Annual Report 1*, 1st June 2008 – 31st May, Salzburg Research Forschungsgesellschaft m.b.h, 2009.

[4] C. M. Tinsley, "The Viking settlement of Northern Iceland". In: R. A. Housley and G. Coles (eds.), *Atlantic Connections and Adaptations: Economies, environments and subsistence in lands bordering the North Atlantic.* Oxford: Oxbow Books, pp. 191-202, 2004.

[5] T. H. McGovern, S. Perdikaris, A. Einarsson and J. Sidell, "Coastal connections, local fishing and sustainable egg harvesting: Patterns of Viking Age inland wild resource use in Myvatn district, Northern Iceland". *Environmental Archaeology,* vol. 11, pp. 187-205, 2006.

[6] T. Amorosi, "Icelandic archaeofauna: A preliminary review". *Acta Archaeologica,* vol. 61, pp. 272-84, 1991.

[7] T. Amorosi, P. Buckland, A. J. Dugmore, J. H. Ingimundarson and T. H. McGovern. "Raiding the landscape: Human impact in the Scandinavian North Atlantic." *Human Ecology,* vol. 25, 3, pp. 491-518, 1997.

[8] http://www.nabohome.org, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 177-181.
ISBN 978-88-8303-295-0. EUT, 2010.

Anthos.es: 10 years showing Spanish plant diversity information in the Internet

Leopoldo Medina, Carlos Aedo

Abstract — Anthos is a scientific and technological program that shows Spanish plant biodiversity information in Internet through the open web page www.anthos.es. The program has been developed since 1999 with the support of the *Fundación Biodiversidad* and the *Agencia Estatal Consejo Superior de Investigaciones Científicas – Real Jardín Botánico* and offers 1.3 million plant data with c. 100.000 annual visitors.

Index Terms — Anthos, Spanish plants, biodiversity, internet.

1 Introduction

Anthos is a scientific and technological program that was developed in accordance with a specific agreement between the *Fundación Biodiversidad* (Biodiversity Foundation, from the Ministry of the Environment) and the *Agencia Estatal Consejo Superior de Investigaciones Científicas* (Spanish National Research Council) – *Real Jardín Botánico* (Royal Botanic Garden, from the Ministry of Science and Innovation) in order to show assorted information about the plant diversity of Spain in the Internet.

The program was initiated in 1999 with a public computer application that has been continually updated up to 1,3 million data concerning plants, using as its main source the Spanish botanical bibliography. In 2005, in accordance with the second agreement for the development of the project, a new computer application was added, developed in a Geographic Information System (GIS), which became accessible to the public in April 2006. A detailed description of the previous Anthos system can be consulted at [1].

The new application integrates and improves the procedures and queries of the previous application, consistently increasing the amount of available data. Furthermore, the new application combines chorological information with other information of a cartographic nature concerning environmental variables, and reference maps. This allows a more accurate location of cited plants, as well as showing in graphic form the different distribution patterns.

The authors are with the Real Jardín Botánico, CSIC, Plaza de Murillo 2, 28014 Madrid (Spain).
E-mail: medina@rjb.csic.es, aedo@rjb.csic.es.

The overall geographical environment chosen for the project is the Iberian Peninsula and the Macaronesian islands (Canaries, Madeira and the Azores) as a representation of each of the biogeographical units present in Spain. In this way, the distribution of a taxon may be studied throughout the entire national territory and surrounding area, fully integrating the taxon in its geographical component.

2 Taxonomic Information

The taxonomic framework for the use and management of the names of plants is still the project *Flora iberica* (www.floraiberica.es), which provides essential knowledge in the areas of taxonomy, chorology, cytology, etc. Anthos follows its criteria faithfully.

Thus, the offered taxonomic treatment is the following:

1. Plants of the Iberian Peninsula and Balearic Islands, within genera which were already published or are in the process of being published in *Flora iberica*. For all other genera, the treatment follows firstly that of Med-Checklist [2] and the rest *Flora Europaea* [3], except in cases where a specific and original treatment is followed.
2. Plants of the Canary Islands, and in general for the entire Macaronesian area, following the taxonomic scheme offered in the *Lista de Especies Silvestres de Canarias* [4].

3 Chorological Information

The distribution maps of plants have been sketched from the chorological information published in scientific articles and books, along with data from herbarium collections, reviewed by specialists who submit their data.

The initial bibliographic information came from the database of chorological citations which the Royal Botanic Garden (CSIC) began to prepare in 1986. This original information was cleaned up and later greatly extended, thanks to the Anthos project, until reaching its current number of 1.3 million entries.

Data from herbarium collections are received from critical reviews normally carried out by authors of genus syntheses for *Flora iberica*, who submit their data to us. In some particular cases, herbarium data for some plants have been added, in order to complete their distribution. The sum of the plant data from herbaria is close to 36,000 sheets.

Recently, we have also added a great amount of duly contrasted information from other databases. This information is shown as it was provided (with the obligatory adaptation for formats), and the source is cited in each record, so that it can be duly identified. We currently have access to databases as the database of “Plantas vasculares de los Parques Nacionales” (Vascular plants of the Spanish National Parks), from the the *Organismo Autónomo de Parques Nacionales*, (Autonomous Organism of National Parks) of the Ministry of the Environment, or the database of the “Plantas vasculares de la cornisa cantábrica” (Vascular plants of the Cantabrian Cornice), submitted by C. Aedo, G. Moreno Moral and Ó. Sánchez Pedraja, members of the group of experts in

botany for the northern part of the Iberian Peninsula, which brings together the great effort this work group has undertaken in the geographical area comprising the Cantabrian Mountains. This database has provided about 300.000 plant references, noticeably improving the quality of the data offered in the area.

Some of the chorological data of plants collected in the botanical bibliography have shown, over time, to be somewhat unreliable. In these cases, although we are obliged to display them to the public, we have marked them with the label "questionable" so that the user may be aware of the fact that the citation needs verification of some type. This label appears in a distinguishable form both on the distribution maps and on the lists.

4 Associated information

Besides the distribution maps for the plants, we have incorporated other information which may be of great interest to users, such as: common names, chromosome numbers, synonymy, conservation state, drawings and photographs. The common names were initially taken from the volumes published in Flora iberica, to which the information contained in the database "*Nombres vernáculos*" (Common names) has been added, gathered and updated by Dr. Ramón Morales (Real Jardín Botánico, CSIC) and his team. This information has been updated in collaboration with Anthos. The information regarding chromosome numbers comes from a previously published database at the *Flora Iberica* webpage, which was subsequently updated with the most recent bibliography. The information on synonyms of the accepted names comes from the database NOMEN of *Flora Iberica*. For the genera not studied in this project yet, the system of nomenclature employed by Med-Checklist and *Flora Europaea* has been adapted to the structure of Anthos, with the aim of obtaining a homogeneous nomenclature database. The information about the conservation state comes from a newly created database in which we have included the Legal Standards on the protection of plants effective in the Spanish territory, together with information on books and red lists. Further information about plant conservation status can be consulted at www.phyteia.es. The illustrations we offer were submitted from several sources: the black and white plates were provided by *Flora Iberica* and were created by different botanical artists. The coloured plates were submitted from other classical works on the Iberian and Macaronesian flora which, due to their antiquity, are no longer subject to authors' or editors' rights. The photographs of the plants were acquired or submitted from diverse artists, whose names appear on the caption of each photo, and which are also responsible for identification of the photographed specimens. In some cases, due to our interest in completing certain collections of images of plants in a geographical or taxonomic area, these photographs were taken within the Anthos project itself, in which case we assume all responsibility for the identification of the plants displayed.

Download Information. Under the epigraph "listings", Anthos has developed a format for the output of data for each consultation on the distribution of a plant. Thus, the user has access to the information that backs up each of the citations. This relation may be downloaded in different formats (txt, csv and xml), which

allows subsequent editing, using the usual geographical and statistical tools.

Cartographic Information. The Cartographic information comes from free public services or was submitted by colleagues.

Google Maps is loaded with the corresponding licenses, as is Blue Marble, and also the climatic variable layers provided by *Atlas Climático de la Península Ibérica* (Climatic Atlas of the Iberian Peninsula, from UAB).

The *Banco de Datos de la Naturaleza* (Nature Data Bank), of the Ministry of the Environment, provided us with the UTM grid, which we later extended to the whole area, with information corresponding to Spanish National Parks.

The information in the Geological Map was taken from the SEIS.NET program, *Sistema Español de Información de Suelos de España sobre Internet* (Spanish Information System of Spanish Soils on the Internet, IRNA-CSIC).

We obtain WMS remote visualisation of the orthophotos of the Rural Register management tool, known as SIGPAC (System of Identification of Agricultural Plots), for which we have availed ourselves of the generous help of those responsible for the above-mentioned application in TRAGSATEC.

The *Instituto Geográfico Nacional* (Spanish National Geographic Institute) suggested and allowed us the use of the WMS service to load layer information provided within the framework of the *IDEE - Infraestructura de Datos Espaciales de España, Ministerio de Fomento* (Spatial Data Infrastructure of Spain, Ministry of Public Works)

The DEM (Digital Elevation Model) was made up by Geodata S.L. from GTOPO30.

5 Use of Information

The information provided in the Anthos Project is distributed on the Internet freely to the broad public for the benefit of whoever may wish to use it; Anthos accepts no responsibility for its reliability, which is the sole responsibility of the authors of the chorological, taxonomic and photographic works.

However, the compilation and management of the above-mentioned information is the work of Anthos, and we should be grateful to be cited as an electronic resource in scientific, technical and professional public outreach works which have availed themselves of the data offered by the program.

Acknowledgements

Since 1999, the year in which Anthos, was initiated, a great number of users have generously collaborated with us, informing us of errors or faults detected or offering information of interest. To all of them, as well as to the institutions and the group of consultants and collaborators both within the *Real Jardín Botánico* or external to it, we owe our deepest thanks for the help which, from the outset and up to this day, we have so generously received. We hope to keep counting on the collaboration of any user, as we are very much aware that this is one of the best ways of correcting and updating the extremely complex information in our pages, and that without the assistance of our collaborators the task would be indeed much more difficult.

References

[1] S. Castroviejo, C. Aedo and L. Medina, "Management of floristic information on the Internet: the Anthos solution". *Willdenowia,* vol. 36, pp. 127-136, 2006.

[2] W. Greuter, H. M. Burdet and G. Long (eds.), *Med-Checklist. A critical inventory of vascular plants of the circum-mediterranean countries 1,3,4*. Conserv. Jard. Bot. Genève, 1984-2008.

[3] S. Castroviejo (ed.), *Flora iberica.* Real Jardín Botánico – Consejo Superior de Investigaciones Científicas. Madrid, 1980-2009.

[4] I. Izquierdo, J. L. Martín, N. Zutita and M. Arechavaleta (eds.), *Lista de especies silvestres de Canarias. Hongos, Plantas y Animales terrestres*. Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias, 2004.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 183-187.
ISBN 978-88-8303-295-0. EUT, 2010.

An interactive tool for the identification of airborne and food fungi

Giovanna Cristina Varese, Antonella Anastasi,
Samuele Voyron, Valeria Filipello Marchisio

Abstract — The growth of fungi may result in several kinds of food-spoilage: off-flavours, discolouration, rotting and formation of pathogenic or allergenic propagules. Moreover many foodborne fungi produce mycotoxins and thus fungal growth in foods and feeds should be avoided. Much interest has also grown for the fungi present in indoor environments, since exposure to airborne biological agents in both the occupational and residential environments could be associated with a wide range of adverse health effects with major public health impact, including infectious diseases, acute toxic effects, allergies and cancer. An interactive identification tool was created for food- and airborne microfungi at the genus and/or species level, based on morphological and physiological data, using the software FRIDA. The interactive key can also be stored on CD- or DVD-roms, or used on media such as PocketPCs of Smartphones. Our key allows the identification of 59 genera/groups and 217 species belonging mainly to Zygomycota and anamorphic and teleomorphic Ascomycota. The database comes with a set of detailed descriptions of each genus and species, a rich archive of images, a glossary of the most frequent mycological terms, and references to descriptions; in addition, culture condition requirements for identification are provided.

Index Terms — Airborne fungi, food spoilage, fungal identification, indoor fungi, interactive keys.

1 Introduction

Today a growing interest of microbiologists is turned to the study of fungal contaminants of food and air. The growth of fungi may result in several kinds of food-spoilage: off-flavours, discolouration, rotting and formation of pathogenic or allergenic propagules. Moreover, many foodborne fungi produce mycotoxins, and thus fungal growth in foods and feeds should be avoided [1], [2], [3]. In the last decades, much interest has also grown for the fungi present

The authors are with the Dipartimento di Biologia Vegetale, Università degli Studi di Torino, viale Mattioli, 25, 10125 Torino, Italia. E-mail of the first author: cristina.varese@unito.it.

in indoor environments, since exposure to airborne biological agents in both the occupational and residential environments could be associated with a wide range of adverse health effects with major public health impact, including infectious diseases, acute toxic effects, allergies and cancer [4], [5].

Although food- and airborne fungi, producing toxins or which cause health hazards, are ubiquitous and belong to the common contamination flora, their recognition is hampered by an incomplete and often confusing literature [1]. Besides, the still poor understanding of many taxonomical groups and the high degree of pleiomorphism in response to environmental changes call for endless floristic, taxonomic and nomenclatural updating.

Moreover, in most of the available books for the identification of fungi the layout follows a hierarchic approach mainly based on classification; hence this approach requires a deep theoretical and practical knowledge of mycology. Even more than for other organisms, therefore, for fungi computer-aided systems are important for the handling of data useful in their identification and as flexible as possible, not necessarily founded on traditional systematic criteria.

We have created an interactive tool for the identification of food- and airborne microfungi at a genus and/or species level. This computer-aided tool can provide access to and simplify the study of fungi by various kinds of users: mycologists, but also those concerned with environmental hygiene (i.e. microbiologists employed in food or pharmaceutical industries), those seeking to create interactive floras, those concerned with the management, planning and conservation of natural resources, and teachers at each educational level.

2 Software

Our interactive identification tool stems from databases created on the basis of morphological, physiological and ecological data of each taxon, using the program FRIDA [6]. Procedures and functions are written in PL/SQL language, running on a Oracle Database engine. FRIDA is flexible, its use does not require the learning of any programming language nor the use of codes to input information and can automatically generate both interactive identification tools, accessible online, and traditional paper-printed identification keys. The keys can be immediately published in the web, and an accessory software was developed to store stand-alone versions on CD- or DVD-ROMs, PDAs (Personal Digital Assistants), and smartphones.

As with most programs for interactive identification, the keys produced by FRIDA are based on a hierarchy of characters, taxa being separated on the basis of those come first in the hierarchy. In our keys, characters are ranked according to the simplicity of observation: macroscopic features of colonies, type of mycelium, presence of ascomata or zygospores, aspect of conidiophores, conidiogenous cells and conidia, etc.

3 The food- and airborne fungi database

The system has produced an interactive key to food- and airborne microfungi which allows the identification of 59 genera and 217 species belonging mainly to

Zygomycota and anamorphic and teleomorphic Ascomycota. The database has detailed descriptions of each genus and species, coupled with a rich pictorial archive of macroscopic and microscopic characters, a brief introduction to the features of the main fungal phyla, explanations of how to cultivate and examine fungi preparing microscopical slides, a glossarium of the more cited mycological terms, and references to descriptions and culture condition requirements.

The interactive keys are usable in two different ways [7]: 1) a simple identification tool based on a traditional dichotomous system, in which the user selects between two options which are explained by means of descriptions, pictures and drawings of the different characters (Fig. 1), 2) a multi-entry query interface in which the user can operate simultaneously a non-hierarchical choice of one or more different characters; FRIDA will select all the taxa with the selected (tagged) characters, and for them it will produce a dichotomous key coupled with pictures of each taxon.

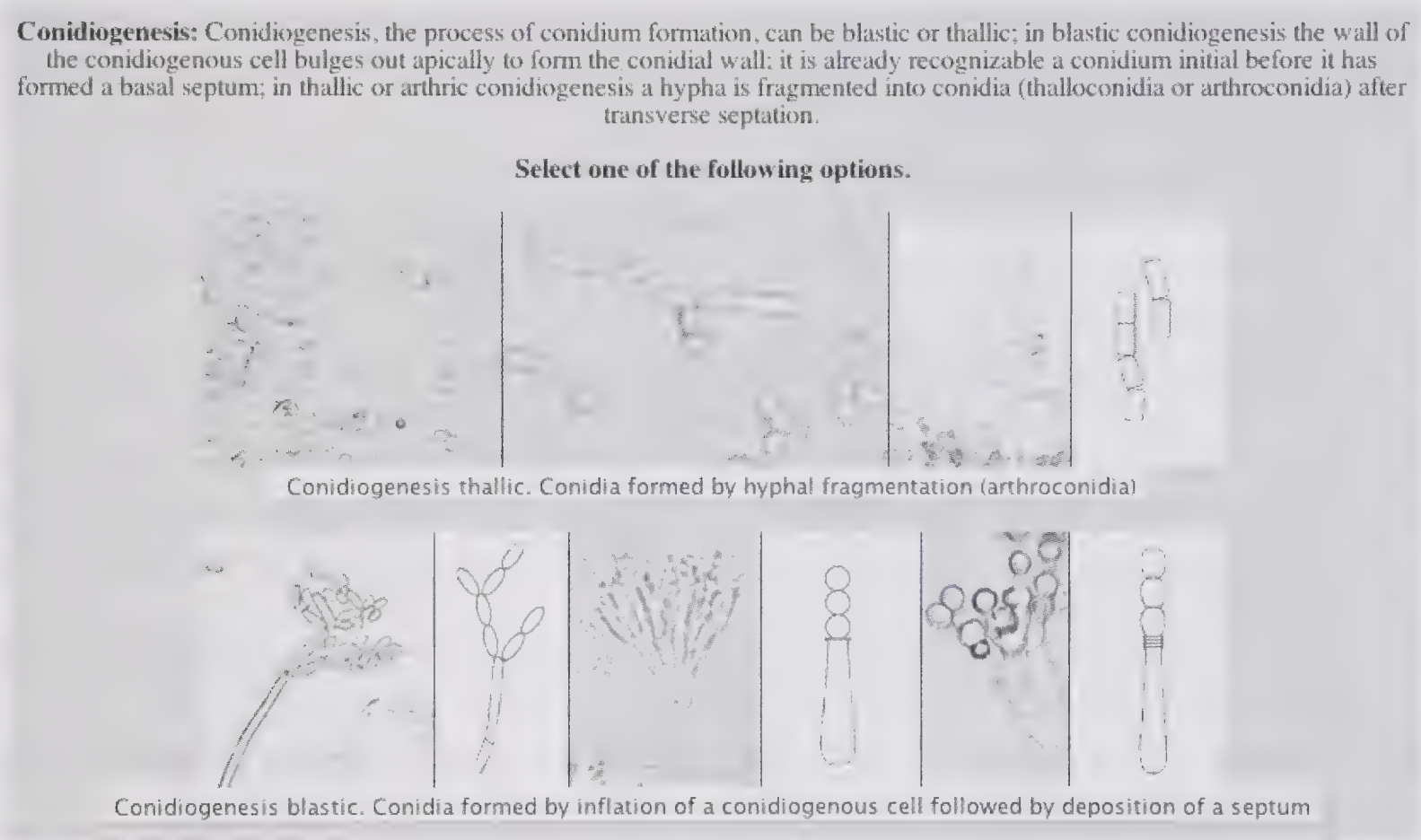

Conidiogenesis: Conidiogenesis, the process of conidium formation, can be blastic or thallic; in blastic conidiogenesis the wall of the conidiogenous cell bulges out apically to form the conidial wall; it is already recognizable a conidium initial before it has formed a basal septum; in thallic or arthric conidiogenesis a hypha is fragmented into conidia (thalloconidia or arthroconidia) after transverse septation.
Select one of the following options.
Conidiogenesis thallic. Conidia formed by hyphal fragmentation (arthroconidia)
Conidiogenesis blastic. Conidia formed by inflation of a conidiogenous cell followed by deposition of a septum

Fig. 1 – Example of a dichotomous key in which the selection between different options is supported by pictures, drawings and descriptions of the most difficult terms.

At the end of an identification process, the system displays a taxon (genus/species) page, reporting the scientific name, a description, and any other information the author has stored in the system (habitat, micotoxins, etc.), as well as an image, and a link to all the images of the same species stored in the image archive. Another important tool of FRIDA is the possibility to insert in the database geographical and ecological data, physiological features, data on the association of a fungus to peculiar substrates, etc. These data can be used to create "filters" which are specific identification pathways encompassing only the species which share the selected character. The "filters" are very useful, since they simplify the identification pathway by reducing the set of species included in a key.

4 Discussion and Conclusions

Since contamination of fodder and foodstuffs and inhalation of propagules suspended in the air exposes people and animals to health risks because of the presence of species producing toxins and MVOC or causing allergies or infections, the use of our key could be useful for biologists working in local health units and similar organizations, as well as in the checking of quality control and environmental hygiene. The prevention of fungi that contaminate indoor environments and cause food spoilage can only be carried out successfully, if the fungal species are known [1]. Knowing the properties of the contaminant species makes it possible to optimize the preservative profile of the food and the hygienic measures in the indoor environments.

However, the identification of these microfungi by means of traditional methods still remains problematic, and exclusively accessible only to a small number of experts. Computer-aided tooks can create a revolution, since they use, in a multi-dimensional way, a wealth of morphological and physiological data, plus the ecological information usually hidden in the large ocean of scientific literature. Traditional keys have several drawbacks that can be avoided by computer-aided tools [7]:

1. Being printed on paper, their content is frozen and hence nomenclatural-taxonomic changes and the discovery of new species render them rapidly outdated. Computerised systems, on the contrary, can be updated and corrected in real time.
2. Traditional keys are rigid. They contain a huge amount of information which is fixed into the format and the logical structure selected by the author. Computerised tools permit to reduce the set of organisms using different combinations of morphological, physiological, ecological, distributional characters i.e. special habitats, mycotoxin production and physiological features (temperature, water activity, pH…).
3. Databases are accumulative. A small database can be the starting point for future expansions.
4. Outputs can be edited in several different formats, from simple texts to illustrated books.

In conclusion, our key, especially if integrated with existing systems based on physiological and molecular criteria, could promote the identification of this important group of organisms even by unskilled persons who lack specific mycological expertise.

Acknowledgement

The authors wish to thank Prof. P. L. Nimis and Dr. S. Martellos for their support in the creation of the database and of the key. This work was supported by a grant from MIUR (Ministero Istruzione, Università e Ricerca).

References

[1] R. A. Samson, E. S. Hoekstra, J. C. Frisvad, *Introduction to food- and airborne fungi,*

Centraalbureau voor Schimmelcultures (CBS), 389 pp., 2004.
[2] C. V. Blackburn, *Food spoilage microorganisms* Woodhead Publishing, 712 pp., 2006.
[3] A. D. Hocking, J. I. Pitt, R. A. Samson and U. Thrane, *Advances in food mycology*, Springer, 371 pp., 2006.
[4] B. Simon-Nobbe, U. Denk, V. Pöll, R. Rid and M. Breitenbach, "The Spectrum of Fungal Allergy" *Int. Arch. Allergy Immunol.,* vol. 145, pp. 58–86, 2008.
[5] J. Brett, J. Green, E. R. Tovey, J. K. Sercombe, F. M. Blachere, D. H. Beezhold and D. Schmechel, "Airborne fungal fragments and allergenicity" *Medical Mycology*, vol. 44, pp. 245-255, 2006.
[6] S. Martellos, "Multi-authored interactive identification keys: The FRIDA (Friendly IdentificAtion) package", *Taxon*, vol. 59 (3), pp. 922-929, 2010.
[7] S. Martellos and P. L. Nimis, "*KeyToNature*: Teaching and learning biodiversity; Dryades, the Italian experience", *Proceedings of the IASK International Conference Teaching and Learning,* pp. 863-868, 2008.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 189-193.
ISBN 978-88-8303-295-0. EUT, 2010.

The Estonian eFlora

Tiina Randlane, Malle Leht, Andres Saag

Abstract — The Estonian eFlora is an example of the new e-learning tools prepared by the *KeyToNature* consortium. It is an interactive digital identification key for more than 1000 plant species recorded from Estonia. The tool is freely available on the internet - in Estonian and English - and has two interfaces (dichotomous and multi-entry), which allow the identification of species using different approaches. Another tool developed by *KeyToNature*, the OpenKeyEditor, allows users not only to edit the text of the existing master key, but also to produce mini-keys restricted to smaller subsets of taxa (e.g. the plants of a park, or a scool garden), and to add user-generated content to them. The reaction to these tools from public media and educational circles has been very positive.

Index Terms — education, Estonia, e-tools, identification, key, plants.

1 Introduction

In the framework of the three-year EU project *KeyToNature* (http://www.keytonature.eu/), where partners from eleven European countries collaborate to produce practical, user-friendly identification tools targeted at the wide audience of teachers and learners, national and/or local keys have been developed for each participating country. Many digital interactive keys for vascular plants and lichens were created using software FRIDA (FRiendly IDentificAtion) and the databases for Italian plants (*Dryades*) and lichens (*Italic*) developed at the University of Trieste, Italy [1], [2].

The new e-learning tools produced by *KeyToNature* have several advantages when compared to the traditional key-books, allowing to widen the users' circle considerably. The new tools can be produced automatically and rapidly by a computer, since the characters are stored in a database. They are created giving more weight to those easy-to-observe characters which make the identification easier to laypersons, such as the colour of flowers, the position and shape of leaves, etc. Once published online, the resulting keys can be updated or edited easily and in real time. Furthermore, there are almost unlimited possibilities to use pictures and drawings to illustrate both the diagnostic characters and the species.

T. Randlane and A. Saag are with the Institute of Ecology and Earth Sciences, University of Tartu, Lai 38/40, Tartu 51005, Estonia. E-mail: tiina.randlane@ut.ee, andres.saag@ut.ee.
M. Leht is with the Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Kreutzwaldi 5, Tartu 51014, Estonia. E-mail: malle.leht@emu.ee.

Estonia is a small North-European country (ca 45 000 km^2, population ca 1,4 million, with a vascular flora of about 1500 species) which was selected as a case study. We wanted to test: 1) whether the database of Italian plants of *Dryades*, including ca 7000 species, is suitable for creating identification keys in other parts of Europe (especially in its northern areas), and 2) whether a digital national-level identification tool would be accepted and practically used by a wide circle of professional and non-professional users, including schoolchildren and students.

2 The Estonian eFlora

2.1 General data

The Estonian eFlora is an interactive digital identification key for ca 1100 plant species (out of ca 1500 native taxa recorded from Estonia [3]). The key also includes, besides indigenous and naturalised taxa, ca. 70 species of introduced trees and shrubs for allowing users to get acquainted with the urban forests. Some taxa which are difficult to separate even for specialists (e.g. species from the genera *Alchemilla, Crataegus, Hieracium, Rosa, Salix, Taraxacum* etc.) are excluded from this key, as well as several very rare species.

The key is currently freely available online in Estonian and in English (http://dbiodbs.univ.trieste.it/carso/chiavi_pub21?sc=368). It has both a dichotomous and a multi-entry interface, which allow identification using different approaches.

2.2 Dichotomous interface

The dichotomous interface is the main instrument of the key, that permits determination of taxa by selecting step by step between two states of a character. For making a selection, one has to click on the corresponding statement-button, after which the next pair of statements is displayed, and so on. A great advantage for beginners is that most character states are richly illustrated by drawings and pictures. At every step of the identification process one can see the number of remaining taxa; clicking on this number, the list of remaining taxa is displayed. At the end of the identification process, the name and a picture of the identified species are shown, and by clicking on the name a taxon page will open.

2.3 Multi-entry interface

In the multi-entry interface the user can choose several characters of a plant in a single step. One just has to specify the characters, click ‘submit’ and wait for a few seconds. The system “filters” the key and gives back a usually strongly reduced list of taxa and – upon request - a smaller dichotomous key for them. This interface can be particularly useful for more expert users, e.g. those who already know the family or the genus of a plant, since it can also produce keys for all species within a family or a genus. It is also the quickest way to go directly to a taxon page by just typing a species’ name.

2.4 Taxon pages

Taxon pages have been compiled for each species to provide important information. The pages in Estonian are more informative than those in English, since they contain short descriptions with the main diagnostic characters, distributional and ecological data, as well as the conservation status of the plant. For most species, distribution maps from the Atlas of the Estonian flora [4] are also displayed. Another attractive feature of the key is that numerous illustrations are available for each species. By clicking on an image, this is strongly magnified, showing even the smallest details. The *Dryades* picture archive presently includes ca. 63.000 pictures of ca. 7300 infrageneric taxa, and is being continuously enriched with new photos and drawings, which in real time become visible also in the online version of the Estonian key.

2.5 OpenKeyEditor

One of the main aims of *KeyToNature* is that of introducing computer-aided identification tools in the educational systems of Europe, including elementary schools. Therefore, one of the main goals is that of rendering our tools as easy as possible. The "difficulty" of a key depends on several factors, such as the selection of characters, the type of interface, the terminology which is used. However, it is obvious that smaller keys (those with fewer species) are generally easier than larger ones. We can now automatically produce identification tools restricted to small subsets of taxa, thanks to the OpenKeyEditor, which we are using also for the Estonian eFlora. This tool allows users to:

1. view and edit the text of the key;
2. add or remove a dichotomy;
3. create smaller keys which are filtered from the master key;
4. automatically generate stand-alone keys for computers and/or smartphones;
5. add user-generated content to the new mini-keys (e.g. adding photos, inserting new text, modifying the terminology, etc.).

Modifying the text of an existing key is easy: it does not require any knowledge of informatic codes or languages, but just the use of a common web browser: one has just to type the changes into the appropriate window. A further function of the OpenKeyEditor permits to create a filter for generating a mini-key (e.g. a key to the plants found in a region of Estonia or in a pond near a school). The filter is just a list of species. To create it, one has just to flag them in a page which lists all taxa included in the Estonian eFlora. The generation of a new mini-key from a filter is easy as well: once the filter is ready, with a single command ("make a key from a filter") the mini-key is generated in a few seconds.

The filtered mini-keys are visible online in real time, since they are produced and hosted by a *KeyToNature* server. However, users may want a stand-alone version of their mini-key. The OpenKeyEditor of *KeyToNature* can produce three different types of stand-alone versions: 1) a CD-Rom version, usable on any computer, 2) a version for PDAs, 3) A version for the i-Phone, which can be disseminated via iTunes.

The use of the OpenKeyEditor for the Estonian eFlora has - for the time being - a restricted access. If you want to use it, please send an e-mail to the Estonian *KeyToNature* contact person (andres.saag@ut.ee).

3 Users' feedback

The Estonian eFlora online was first presented to a wide audience of students, teachers and citizens in Tartu (September 2009), on the occasion of a yearly meeting of *KeyToNature*. The presentation was broadly and positively reflected in the national media [5], [6], [7], contributing considerably to the progress of public interest. In the first week only, more than 1700 users (or just watchers) visited our site.

The first part of the eFlora, limited to woody plants, was available online much earlier, with applications for iPhone, iPodTouch and iPad from iTunes [8], which fostered a great interest of media.

The Estonian eFlora was mainly created for teachers and their students. Teachers - from school teachers to university professors - feedbacked us through questionnaires on their experience with the eFlora. Altogether, 19 using events with about 350 participants have been officially recorded in Estonia on three different educational levels (primary and secondary schools, and universities). However, the actual number of users in school lessons is probably much higher, as not all teachers have filled the questionnaire. According to the answers in the questionnaire, the computer-based activities for identifying organisms are very much appreciated by both teachers and students (based on teachers' judgements). The huge amount of images connected to the keys was seen as a primary positive aspect. Problems with scientific terms used in the keys but not understood by pupils occurred especially in the younger classes. Several teachers had solved the problem by preparing an introductory part to the lesson, during which specific terms were explained. As the identification practices were much acknowledged, it has been often proposed to include these activities in the context of the schools' official curriculum.

Acknowledgement

The Estonian eFlora and its OpenKeyEditor were prepared within the project *KeyToNature* in cooperation between the University of Trieste (Italy), the University of Tartu and the Estonian University of Life Sciences (Estonia). *KeyToNature* is financed by the European Union through the program *e*Content*plus*. Special thanks are due to Aino Kalda, Thea Kull and Ülle Reier who provided important input. Rein Kalamees, Jaan Liira, Jaanus Paal, Kersti Püssa, Elle Roosaluste, Kai Rünk and Tiina Talve are acknowledged for putting several plant pictures at our disposal.

References

[1] S. Martellos and P. L. Nimis, "*KeyToNature*: Teaching and Learning Biodiversity: Dryades, the Italian Experience". In: M. Muñoz, I. Jelinek and F. Ferreira (eds.), *Proceedings of the International Association for the Scientific Knowledge (IASK) International Conference "Teaching and Learning"*, pp. 863–868, 2008.

[2] T. Randlane, A. Saag, S. Martellos and P. L. Nimis, "Computer-aided, interactive keys to

lichens in the EU project *KeyToNature*, and related resources". In: T. H. Nash III (ed.), *The lives of the lichen symbionts*, Bibliotheca Lichenologica, vol. 105, Stuttgart, J. Cramer in der Gebrüder Borntraeger Verlagsbuchhandlung, (in press), 2010.

[3] T. Kukk, *Eesti taimestik*. Tartu–Tallinn: Teaduste Akadeemia Kirjastus, 464 pp., 1999.

[4] T. Kukk and T. Kull (eds.), *Atlas of the Estonian flora*. Institute of Agricultural and Environmental Sciences of the Estonian University of Life Sciences, Tartu, 527 pp., 2005.

[5] T. Randlane, "An interview in a national radio broadcast", *Environmental Tent*, available at http://www.eseis.ut.ee/k2n_promo/keskkonnatelk20090222_1.mp3, 22.02.2009.

[6] U. Käärt, "Eesti e-taimetark internetis meelitab magnetina tuhandeid loodusesõpru", *Eesti Päevaleht*, Available at http://www.epl.ee/artikkel/480088,13.10.2009.

[7] T. Randlane, "Digitaalsed taime- ja samblikumäärajad ootavad koolides laiemat kasutamist", *Koolielu,* Available at: http://arhiiv.koolielu.ee/pages.php/0710,24405, 25.11.2009.

[8] A. Saag, T. Randlane and M. Leht, "Keys to plants and lichens on smartphones – Estonian examples". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 195-199, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 195-199.
ISBN 978-88-8303-295-0. EUT, 2010.

Keys to plants and lichens on smartphones: Estonian examples

Andres Saag, Tiina Randlane, Malle Leht

Abstract — The EU project *KeyToNature* aims to contribute to a better knowledge of biodiversity by a practical activity, the identification of species. The project is introducing new tools, including digital keys for mobile media. In Estonia two applications have been prepared for smartphones: a 'Key to trees and shrubs of Estonia' (for iPhone, iPodTouch and iPad) and a 'Key to Estonian epiphytic macrolichens' (using the Android operating system). A third tool, the 'Key for plants of the island Naissaar' is an example of application of the Open Key Editor for creating new keys for mobile devices, starting from a larger master key.

Index Terms — Estonia, e-tools, identification, keys, smartphones.

1 Introduction

The UN has declared 2010 to be the International Year of Biodiversity. The world is invited to take action to safeguard the variety of life on Earth. Protecting or managing in a sustainable way the biota in any part of the world is possible only when their species are not only recorded and recognized by specialists, but also noticed and appreciated by the widest of audiences – by everyone.

The three-year EU project *KeyToNature* (http://www.keytonature.eu/), which started in 2007 and is finishing in autumn 2010, aims to contribute to a better knowledge of biodiversity by a practical activity, the identification of species.

The consortium introduces new tools for this purpose, keys that allow the easy identification of plants, lichens, birds and other organisms using digital, user-friendly facilities [1]. Hundreds of digital keys have been produced within *KeyToNature*, many of them in close cooperation with the University of Trieste, Italy, using FRIDA (FRiendly IDentificAtion), a software developed in this University. The keys are published online and are freely accessible to

A. Saag and T. Randlane are with the Institute of Ecology and Earth Sciences, University of Tartu, Lai 38/40, Tartu 51005, Estonia. E-mail: andres.saag@ut.ee, tiina.randlane@ut.ee.
M. Leht is with the Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Kreutzwaldi 5, Tartu 51014, Estonia. E-mail: malle.leht@emu.ee.

everyone. They can also be stored on CD-ROMs, to be used without an Internet connection. Those who still prefer paper-printed keys can print out their own "keybook". The latest trend is to develop applications which permit the use of digital keys on mobile media, such as palmtop computers and smartphones, either online or in stand-alone form.

2 Digital keys on mobile media

2.1 General

A smartphone is a mobile phone that offers more advanced computing ability and connectivity than a contemporary basic mobile phone. Smartphones allow the user to install and run various applications based on a specific platform, as they run complete operating system software providing a platform for application developers. These advanced mobile devices possess powerful processors, abundant memory, larger multi-touch screen and a virtual keyboard with e-mail, web browsing and WiFi connectivity. Today smartphones form the fastest growing segment of the mobile phone market [2].

Identification keys in the mobile phones show several advantages compared to the usual digital keys accessible through Internet or on CD-ROMs. Firstly, they enable the use of identification tools not merely at home, in a classroom or in a lab, but also in the field. The possibility to identify the exact geographical location of a plant by GPS (inserted in the mobile phone) and to send additional information (photos, notes, descriptions) via email – for example, to a botany expert for the evaluation of an identification – is a valuable supplementary function of such applications.

2.2 A botanist in your pocket: the key to Estonian trees and shrubs

The Estonian eFlora is an interactive digital identification key for c. 1100 plant species (out of ca 1500 taxa recorded from Estonia), including several introduced tree species. Some taxa from critical genera are excluded from this key, as well as several very rare species. The key was presented to a wide audience of students, teachers etc. in Tartu (Estonia), in September 2009, and is permanently freely available in Estonian and in English at http://dbiodbs.univ.trieste.it/carso/chiavi_pub21?sc=368.

Presently, two adaptions for smartphones are available: a 'Key for trees and shrubs of Estonia' (for iPhone, iPodTouch and iPad), and a 'Key for plants of the island Naissaar' (using the Android operating system).

The application 'Key for trees and shrubs of Estonia' (Fig. 1) allows users to:

1. Identify more than 140 tree and shrub species in Estonia.
2. See the explanatory files about species and access the photo gallery.
3. Search by taxon name.
4. Post the identified species to Facebook.
5. Take pictures and add field notes to one's guide.
6. Send multimedia content (pictures and notes) stored in the device via email.

The application was developed by Divulgando Srl (Italy) and released publicly at the end of 2009. It is available for download from the iTunes App Store for a symbolic price of 2,39 EUR. During a short period (18.–24. January 2010), it was even at the very top of the list of Paid Apps in the iTunes App Store and still today it is top ranking in those for the educational sector.

The 'Key for plants of the island Naissaar' is an example of a filtered key derived from the Estonian eFlora, generated by a new sofware, the OpenKeyEditor of *KeyToNature* [3] specifically for a mobile device. This key, which includes 415 plant species, was ordered by a company carrying out nature-educational training courses on the island Naissaar.

Both tools display a dichotomous interface where each step of the identification process is richly illustrated with pictures and drawings [3]. As the applications can be downloaded to the memory card of a smartphone, they can be used in stand-alone form without additional web-browsing charges (with some limitations in the access to the image archives, compared to the online keys).

2.3 A lichenologist in your pocket: the key to Estonian epiphytic macrolichens

Another application for smartphones, using the Android operating system, has been developed by the Company Mine Avasta (Estonia), based on an internet key which was produced in cooperation between the University of Trieste (Italy) and the University of Tartu. This application enables the identification of 115 species of epiphytic macrolichens known to occur in Estonia (Fig. 1). The main principles are similar to those of the previously described application: it allows identification of taxa using a simple dichotomous key in which the user has to decide between two options; it is also possible to search the taxon by its Estonian or Latin name, and then get additional information about the species – read a summary of diagnostic characters, and see the photos and distribution map of the species in Estonia. As the characters of lichens which are used in the key are less familiar to the wide audience, an explanation of the main characters of lichens in the form of an illustrated glossary is also provided.

The tool was uploaded into the Android Market in July 2010, and is free of charge.

3 Conclusion

The two Estonian examples of digital identification keys for smartphones were meant to attract the attention of a wide circle of non-specialists: pupils, students, teachers, forestry workers, nature conservation staff, tourists etc. – to increase public awareness of biodiversity and to allow new approaches in nature education. Both tools are available not only as smartphone applications but also on the Internet as interactive keys, freely accessible to everyone (http://dbiodbs.univ.trieste.it/carso/chiavi_pub21?sc=175, trees and shrubs; http://dbiodbs.univ.trieste.it/carso/chiavi_pub21?sc=159, epiphytic macrolichens).

We have already received positive feedback from a large audience through

Fig. 1 – Front pages of the mobile applications 'Key for trees and shrubs of Estonia' (left) and 'Key for the identification of Estonian epiphytic macrolichens' (right).

public media (newspaper articles, electronic publications, broadcasts) indicating that the local society is willing to accept and use the new interactive e-learning devices in everyday life [4], [5], [6].

Acknowledgement

The described facilities have been prepared within the project *KeyToNature,* financed by the European Union through the program *e*Content*plus* as well as through the European Regional Development Fund (Center of Excellence FIBIR). Special thanks are due to Rodolfo Riccamboni (Divulgando Srl, Italy) and Marko Peterson (Mine Avasta, Estonia) for developing the mobile applications.

References

[1] S. Martellos and P. L. Nimis, "*KeyToNature*: Teaching and Learning Biodiversity: Dryades, the Italian Experience". In: M. Muñoz, I. Jelinek and F. Ferreira (eds.), *Proceedings of the International Association for the Scientific Knowledge (IASK) International Conference "Teaching and Learning"*, pp. 863–868, 2008.

[2] "Smartphone definition from PC Magazine Encyclopedia". PC Magazine, available at http://www.pcmag.com/encyclopedia_term/0,2542,t=Smartphone&i=51537,00.asp, July 2010.

[3] T. Randlane, A. Saag, S. Martellos and P. L. Nimis, "Computer-aided, interactive keys to lichens in the EU project *KeyToNature*, and related resources". In: T.H. Nash III (ed.), *Together and separate: The lives of the lichen symbionts*, Bibliotheca Lichenologica, vol. 105, Stuttgart,

J. Cramer in der Gebrüder Borntraeger Verlagsbuchhandlung, (in press), 2010.
[4] M. Aeltermann, “Eesti e-Floora määraja võimaldab tuvastada taimeliike”, *ERR Uudised,* avalable at http://uudised.err.ee/index.php?06191611, 18 January 2010.
[5] U. Käärt “Telefon määrab teadlaste abiga puu- ja põõsaliike”, *Eesti Päevaleht*, available at http://www.epl.ee/artikkel/486512, 19 January 2010.
[6] M. Himma, “Taimed näitavad end nutitelefonis”, *Tartu Postimees,* available at http://www.tartupostimees.ee/?id=214092, 20 January 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 201-205.
ISBN 978-88-8303-295-0. EUT, 2010.

IIKC: An Interactive Identification Key for female *Culicoides* (Diptera: Ceratopogonidae) from the West Palearctic region

Bruno Mathieu, Catherine Cêtre-Sossah, Claire Garros, David Chavernac, Thomas Balenghien, Régine Vignes Lebbe, Visotheary Ung, Ermanno Candolfi, Jean-Claude Delécolle

Abstract — In 2006, bluetongue virus (BTV) outbreaks appeared surprisingly in northern Europe and widely affected most of the European countries. Correct identification of *Culicoides* species (Diptera: Ceratopogonidae), known as BTV vectors, is a key component of all studies intending to understand vector dynamics and to develop vector control strategies. A computer-based system, Xper^2, was used to develop an Interactive Identification Key (IIKC) for female *Culicoides* from the West Palearctic region. The current version of IIKC includes 108 taxa, 61 descriptors and 837 pictures and schemes. IIKC is a powerful tool for routinely identifying *Culicoides* species and for training young specialized taxonomists.

Index Terms — *Culicoides* species, identification key, interactive key, bluetongue.

1 Introduction

Bluetongue is an arboviral disease affecting ruminants, mainly ovines. Vectors of bluetongue virus (BTV) are small biting midges belonging to the genus *Culicoides* (Diptera: Ceratopogonidae). Excluding the 46 fossils, a total of 1308 *Culicoides* species are distributed on every large land mass with the exception of Antarctica and New Zealand, ranging from the tropics to the tundra and from sea level to 4000 m [1]. Worldwide, around 60 biting midges species are suspected or proved to transmit viruses, protozoa or filaria worms

B. Mathieu, E. Candolfi and J.-C. Delécolle are with the Institut de Parasitologie et de Pathologie Tropicale, Université de Strasbourg, EA 4438, 67000 Strasbourg, France. E-mail: bmathieu@unistra.fr.
C. Cêtre-Sossah, C. Garros, D. Chavernac and T. Balenghien are with Cirad, UMR Contrôle des maladies animales exotiques et émergentes, F-34398 Montpellier, France.
R. Vignes Lebbe and V. Ung are with UMR 7207 CNRS-MNHN-Université Pierre et Marie Curie, 75005 Paris, France.

[2]. Several of these viruses are of major international significance for animal health (African horse sickness virus, bluetongue virus, epizootic hemorrhagic disease virus).

Up to 2010, five autochthonous biting midges species are suspected to transmit BTV in western and central Europe: *Culicoides chiopterus* (Meigen, 1830) [3], *C. dewulfi* Goetghebuer, 1935 [4], *C. obsoletus* (Meigen, 1919) [5], *C. scoticus* Downes and Kettle, 1952 [6] and *C. pulicaris* (Linnaeus 1758) [7]. In addition, along the Mediterranean basin, the main afro-tropical vector *Culicoides imicola* Kieffer, 1913 is present and is in charge of BTV transmission. The emergence and spread of bluetongue disease in Western Europe has highlighted the taxonomic impediment concerning biting midges. Identification of *Culicoides* species is highly important for a clear understanding of virus transmission and biting midges population dynamics and for surveillance activities as well. Because of their small size and of their highly specific diversity, morphological identification of biting midges requires time and expertise.

Despite of the fact that monitoring activities are compulsory, most of the European countries affected by bluetongue virus have very few taxonomists for biting midges. They rely on two main identification keys: Campbell and Pelham-Clinton's key, published in 1960 [8], and Delécolle's key, published in 1985 [9]. These two dichotomous keys are remarkable, but they have not been updated for years and generally, they are not well adapted for non-expert researchers. Therefore, when using the keys, newly described species or synonymies can be missed, and scientists face difficulties with diagnostic and described characters.

Since a decade, the development of computer-aided systems marks a turning point in taxonomy [10; 11]. Interactive identification keys based on multi-entries are easy to use for experts and non-experts. They allow quick updates and are easily released to the scientific community through the web. Today, interactive identification keys have been developed for several arthropods: phlebotomine sandflies [12], *Glossina* flies [13], or mosquitoes [14].

The aim of this work is to present the newly developed interactive identification key for female *Culicoides* for the West Palearctic region (IIKC). Information on availability and some recommendations are given.

2 Material and methods

2.1 Morphological characters

60 morphological descriptors coded in 164 morphological states are observed from wings, abdomen, head and leg.

The wings descriptors are related to the presence or absence of spots, position, size and shape.

Abdominal descriptors concern the number, size and shapes of spermathecae, the presence and shape of a sclerotized ring, and some special features like abdominal sclerites.

The head descriptors gather those of 4 body parts: antennae, eyes, palps and mouthparts. Antennae are observed for the sensilli distribution (coeloconica,

short and long trichodea) and the antennal ratio (length of the first elongated segment divided by the last short one). On eyes, inter-ocular space and the interfacetal hairs are observed. Data related to the shape of the palpal segments and the sensory pits are collected. Teeth are observed on mandibles and maxilles.

The morphological characters were discussed and validated by 27 entomologists from 14 countries at the taxonomy meeting of the MedReoNet network at Strasbourg in 2009 (http://medreonet.cirad.fr/).

2.2 Software and collections

The morphological database of female *Culicoides* was edited with Xper^2 version 2.0 [15] which allows the creation of the interactive key.

Slide-mounted specimens observed are from the Callot, Kremer and Delécolle collections (IPPTS, Strasbourg, France). Pictures were taken with a Zeiss® microscope equipped with a Motic® camera. All pictures were individually cleaned up using the software Gimp version 2.6.2.

2.3 Validation

A trial has been proposed to 5 non expert scientists on *Culicoides* taxonomy. 37 slide-mounted female specimens belonging to the genus *Culicoides* are anonymously coded. These selected species span a wide range of morphological diversity of *Culicoides*. The identification order of the slides has been randomly determined for each participant. Correct or false status of identification, estimates of user's confidence and the use of definitions and illustrations were also gathered.

3 Results

In total, 60 morphological descriptors have been observed: 27 on the wing, 14 on the abdomen, 16 on the head and 3 on the legs. An additional geographical descriptor has been added, which allows users to limit the taxa list to one country. The 60 morphological descriptors are divided into 164 morphological states illustrated with 403 pictures and schemes. Morphological data of 22 species were collected from stocked species at IPPTS, Strasbourg, France and, 86 others species were from Callot, Kremer and Delécolle collections. In total, the current version of IIKC includes 108 taxa. Among them, 8 species with important morphological variations have been coded as taxa with polymorphic characters. A total of 76 taxa were illustrated with drawings sheets. 73 species were illustrated with 434 pictures (mean of 5.9 pictures per taxon): 24 with only pictures and 49 with both pictures and drawings. Only 8 taxa have not yet been illustrated. IIKC includes a total of 837 pictures and schemes.

IIKC is still in a validation step at the submission date of this communication thus, results could not be shown.

4 Discussion

4.1 Availability and updates

IIKC will be freely available on a CD-Rom upon request to the authors. Moreover, a dedicated website under construction will allow the interactive key as well as updates to be downloaded. To help users with limited computer capacity, IIKC will also be available on-line without local installation.

A scientific committee has been proposed, and annual meetings will be organized to validate updates, discuss new species or synonymies, evaluate new systematic or taxonomic changes. IIKC users are encouraged to contact the authors and the scientific committee for feedbacks and to inform for new taxonomic information.

4.2 Recommendations

IIKC helps in identifying adult female *Culicoides* species. Identification of *Culicoides* genus among other genus of the Ceratopogonidae family is not included.

IIKC helps in identifying slide-mounted specimens and users are recommended to use a microscope for good morphological observations. Stereomicroscope observations of biting midges preserved in alcohol limit the observation on wing patterns only.

IIKC is a multi-entry key. Compared to a dichotomous key, this key allows to choose the descriptors the user wants to observe. If the specimen has damaged parts, identification can go on with others ones. Users can also select a group of descriptors (wing, abdomen, leg, head, or geography). Three optimized list of characters classifying descriptors according to their discriminating power, are available as an option and leads to quick identification. Users are strongly recommended to use the option “Xper original sort”. When activated, a number into brackets appears for each descriptor (from 0 to 1) representing the discriminating power. The highest numbers are the most powerful descriptors (i.e. the ones that will best discriminate the taxa).

Users are strongly recommended to use reference collections and national experts to confirm their identification when dealing with new recorded or observed species. Authors strongly encouraged future users to build regional reference collections and to help in exchanging material between collections to improve our systematic and taxonomic knowledge of the genus *Culicoides*.

5 Conclusion

IIKC is a newly developed morphological identification key allowing the identification of 108 taxa of *Culicoides* (Diptera: Ceratopogonidae). Largely illustrated with 837 pictures, drawings and schemes, this interactive identification key is based on a multi-entry system, with optimized list of characters (including geographical distribution). The richness of illustrations is a great advantage to train taxonomists. The development of identification tools for *Culicoides* and more

generally for arthropods involved in pathogen transmission will help scientists in identifying species and therefore will give better insights into the bioecology and dynamics of these groups, helping in designing more appropriate vector control strategies.

Acknowledgements

IIKC was developed in the framework of a surveillance network of reoviruses, bluetongue and African horse sickness in Mediterranean basin and Europe (acronym MedReoNet, contract no. SSPE-CT-2006-044285. The authors would like to acknowledge all the partners of the WP 2, partners of MedReoNet, and especially R. Meiswinkel, F. Schaffner, M. Miranda. Special thanks to X. Allène, S. Carpenter, D. Delécolle, L. Gardès, K. Huber, I. Rakotoarivony, M.L. Setier-Rio, R. Vénail for testing the beta version and the validation step.

References

[1] A. Borkent, *World species of biting midges (Diptera: Ceratopogonidae),* Belmont University, The Ceratopogonid web page, pp. 236, 2009.

[2] A. Borkent, "Chapter 10. The biting midges - The Ceratopogonidae (Diptera)", in: W.C. Marquardt, (ed.), *Biology and disease vectors*, Elsevier Academic Press, pp. 113-126, 2005.

[3] E. Dijkstra, I. J. van der Ven, R. Meiswinkel, D. R. Holzel and P. A. Van Rijn, "*Culicoides chiopterus* as a potential vector of bluetongue virus in Europe", *Vet. Rec.*, vol. 162, p. 422, 2008.

[4] A. Stephan, P. H. Clausen, B. Bauer and S. Steuber, "PCR identification of *Culicoides dewulfi* midges (Diptera: Ceratopogonidae), potential vectors of bluetongue in Germany", *Parasitol. Res.*, vol. 105, pp. 367-371, 2009.

[5] S. Carpenter, H. L. Lunt, D. Arav, G. J. Venter and P. S. Mellor, "Oral susceptibility to bluetongue virus of *Culicoides* (Diptera: Ceratopogonidae) from the United Kingdom", *J. Med. Entomol.*, vol. 43, pp. 73-78, 2006.

[6] S. Carpenter, C. McArthur, R. Selby, R. Ward, D. V. Nolan, A. J. Luntz, J. F. Dallas, F. Tripet and P. S. Mellor, "Experimental infection studies of UK *Culicoides* species midges with bluetongue virus serotypes 8 and 9", *Vet. Rec.*, vol. 163, pp. 589-592, 2008.

[7] S. Caracappa, A. Torina, A. Guercio, F. Vitale, A. Calabro, G. Purpari, V. Ferrantelli, M. Vitale and P. S. Mellor, "Identification of a novel bluetongue virus vector species of *Culicoides* in Sicily", *Vet. Rec.*, vol. 153, pp. 71-74, 2003.

[8] J. A. Campbell and E. C. Pelham-Clinton, "A taxonomic review of the british species of *Culicoides* Latreille (Diptera: Ceratopogonidae)", *Proc. R. Soc. Edinburgh*, vol. 68, pp. 181-302, 1960.

[9] J. C. Delécolle, *Nouvelle contribution à l'étude systématique et iconographique des espèces du genre* Culicoides *(Diptera: Ceratopogonidae) du Nord-Est de la France*, PhD dissertation, U.F.R. sciences de la vie et de la terre, Université Louis Pasteur de Strasbourg I, pp. 229, 1985.

[10] D. Agosti, "Biodiversity data are out of local taxonomists' reach", *Nature*, vol. 439, p. 392, 2006.

[11] D. E. Walter and S. Winterton, "Keys and the crisis in taxonomy: extinction or reinvention?", *Annu. Rev. Entomol.*, vol. 52, pp. 193-208, 2007.

[12] R. Vignes Lebbe and C. Gallut, *Computer Aided Identification of Phlebotomine sandflies of Americas (CIPA)*, Université Pierre et Marie Curie, Paris, France. http://lis-upmc.snv.jussieu.fr/xper2/infosXper2Bases/en/, 1997.

[13] J. Brunhes, D. Cuisance, B. Geoffroy and J. Hervy, *Les glossines ou mouches tsé-tsé (réédition)*, IRD Editions, Montpellier, France (CD-Rom), 2009.

[14] F. Schaffner, G. Angel, B. Geoffroy, J. Hervy, A. Rhaiem and J. Brunhes, *The mosquitoes of Europe. An identification and training software*, IRD Editions and EID Méditerranée, Montpellier, France (CD-Rom), 2001.

[15] V. Ung, G. Dubus, R. Zaragueta-Bagils and R. Vignes Lebbe, *Xper²: introducing e-taxonomy*, *Bioinformatics*, vol. 26, pp. 703-704, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 207-211.
ISBN 978-88-8303-295-0. EUT, 2010.

Indochinese bamboos: biodiversity informatics to assist the identification of "vernacular taxa"

My Hanh Diep Thi, Régine Vignes Lebbe, Ha Phuong Nguyen, Bich Loan Nguyen Thi

Abstract — Bamboo (Bambusoides – Gramineae) is one of the most important natural resource in Southeast Asia. However, bamboo identification has many difficulties. In the area of SEP programme «Indochina Bamboos» (2008-2010), the final objective is to update the bamboo flora of Indochine (Vietnam, Laos, Cambodege) and to publish an e-flora including free access keys, digital images and information about the traditional and economic use of bamboos. During field trips in Vietnam, Laos and Cambodgia, samples, pictures and morphological description sheets were collected in various locations and previously assigned to the local vernacular names. We use an informatic program, Xper^2, to assist the comparison and identification of "vernacular bamboo taxa" based on morphological characteristics.

Index Terms — bamboo, Indochinese, taxonomy, common name, computing, XPER^2.

1 Introduction

Bamboo is extensively used in traditional handicrafts in Southeast Asia. Nowadays, bamboo is also used in others fields: construction, medicine, etc. With a better understanding of this group of plants, we could propose recommendations on conservation measures and find species to be developed for industrial exploitation and economic benefits [4].

Bamboo is in the family Gramineae (Poaceae), subfamily Bambusoideae, tribe Bambuseae. Since Linné time, its taxonomy is based on flower characteristics. However, bamboo is characterized by infrequent flowering. The taxonomy

M. H. Diep, is with the University of of Sciences of HCMC, Centre de Recherche pour la Conservation des ressources naturelles (CRC). Vietnam and Director of the Phu An Plant Conservation Centre, Vietnam.
R. Vignes Lebbe is with UMR 7207 CNRS/MNHN/UPMC, MNHN Département Histoire de la Terre, CP48, 57 rue Cuvier, 75005 Paris, France.
H. P. Nguyen is a master student in the Université Pierre et Marie Curie, Paris –VI, France
B. L. Nguyen Thi is master student in the University of of Sciences of HCMC, Vietnam.

of Indochinese Bamboo has not been completed; it is basically based on E. G. Camus and A. Camus, «Flore Générale d'Indochine, Vol 7 Gramineae» [9] describing 14 genera and 73 species [6]. In the 1970s, Professor Pham Hoang Ho mentioned more than 120 bamboo species in «Vietnamese Plants» [8]; almost 200 species with illustrative pictures are recorded in Nguyen Hoang Nghia, «Vietnamese Bamboos» [7].

Facing the ongoing disappearance of many traditional uses of bamboo, and its shrinking natural environment, Dr. Diep Thi My Hanh has decided to establish the Bamboo Ecology Museum and the Plant Conservation Centre in Phu An, to collect a variety of bamboo species and other endangered precious plants in the Southeast. The project is jointly undertaken by Rhône-Alpes (France), the Binh Duong Province (Vietnam), the Pilat Natural Garden, and the Natural Science University of HCMC. In 2003-2007, the project has gathered a large amount of information on Vietnamese bamboos in the North, Central Part, Highlands, Mekong Delta, and the Southeast, with 301 dry specimens in a botany collection and 157 samples of bamboos growing in the Conservation Centre [2].

Since 2007, a project to achieve the revision of Indochine Bamboos is in progress, in collaboration with Laotian, Cambodgian and Vietnamian biologists. During many field trips, morphological description sheets with pictures and information on the bamboo applications in various locations have been made. With a few exceptions, most samples have their common names in each location. A crucial task is then to assign a scientific name to all gathered data. This paper describes the methodology and the results of the project.

2. Methodology

2.1 Data collection

Data were gathered from literature, collections and field trips. Field trips were conducted for the most part in Vietnam (all the regions) and in Laos and Cambodia as well. The exploration needs to be completed in some locations. The literature was consulted and analysed to collect all characters proposed by botanists to define and identify bamboo species. This task was completed by the observation of specimens (including type material) in the main reference collections, such as the Royal Botanic Gardens Kew (UK), the Laboratoire de Phanérogamie de Paris (France), and other botany collections in Asia.

2.2 Proposal of a standardised description form to describe specimens and Bamboo species

A list of 90 morphological characters, divided into 11 groups, has been established and documented by texts and images (Fig. 1). The botanic terminology was controlled by the botanist Soejatmi Dransfield. The database is now translated into five languages (Vietnamese, English, French, Laotian, and Cambodian).

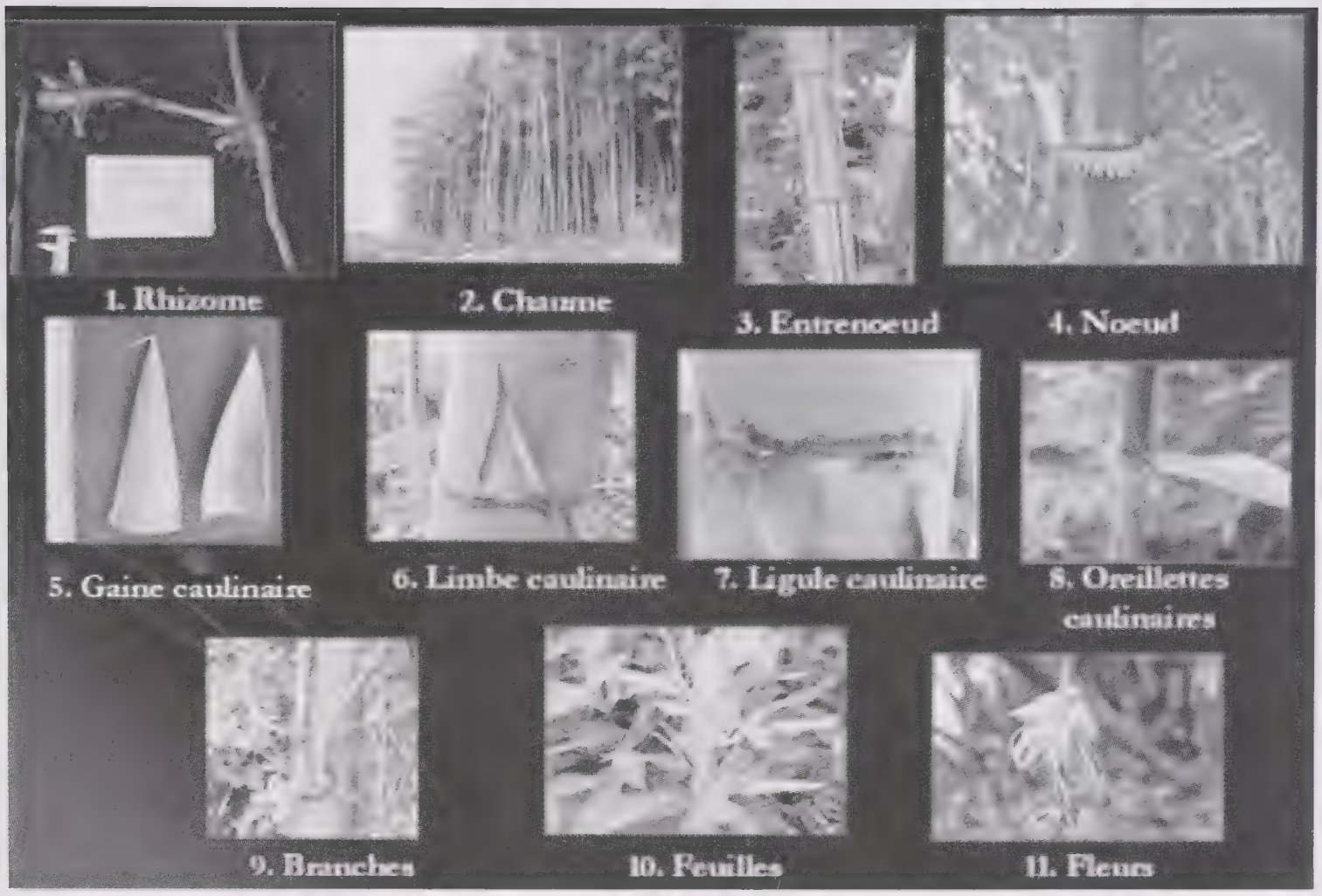


Fig. 1 – The 11 groups of characters describing Bamboos.

2.2 Digitalization of the Bamboo descriptions

Xper^2 appears well adapted to manage our structured descriptions, texts and images. Following the standardized list of characters, we edit the descriptions of the species and also the descriptions of specimens with their common names.

2.3 Assigning scientific names to vernacular names

The scientific identification of each specimen is time-consuming and requires highly skilled and adequately trained scientific personnel. To facilitate this task, we use the facilities offered by Xper^2 to compare descriptions (see Fig. 2). We also use the free access key of Xper^2 to associate the specimens of the references collections to “vernacular taxa”. Similarity measurements between descriptions are used to group vernacular and scientific descriptions. All these results are compared, to propose one or few scientific names for each vernacular name. This work is already in progress.

2.4 DNA analyses

To complete and to verify some identifications, DNA analyses are conducted iin collaboration with the MNHN and the laboratory of Créteil University. This early approach enables us to conduct more intensive studies on the Indochinese Bamboo evolution based on analyzing the molecular evolution.

Legend
Discrimination
Partial discrimination
No discrimination

	Dendrocalamus giganteus Munro	Tre mỡ gai (Hà Tĩnh)-C	UNION	INTERSECTION
Type de rhizome (Dạng căn hành) (Type of rhizome)	sympodial (cespiteux) (cộng trụ) (sympodial)	sympodial (cespiteux) (cộng trụ) (sympodial)	sympodial (cespiteux) (cộng trụ) (sympodial)	sympodial (cespiteux) (cộng trụ) (sympodial)
Taille du chaume (Chiều cao thân) (Size of culm)	géant: supérieure à 20 m (rất lớn: lớn hơn 20 m) (giant : greater than 20 m)	grand: 7 à 20 m (lớn: 7đến 20 m) (large : 7 to 20 m)	géant: supérieure à 20 m (rất lớn: lớn hơn 20 m) (giant : greater than 20 m), grand: 7 à 20 m (lớn: 7đến 20 m) (large : 7 to 20 m)	
Diamètre du chaume (Đường kính thân) (Culm diameter)	10 à 20 cm (10đến 20 cm) (10 to 20 cm); supérieur à 20 cm (lớn hơn 20 cm) (greater than 20 cm)	5 à 10 cm (5đến 10 cm) (5 to 10 cm)	10 à 20 cm (10đến 20 cm) (10 to 20 cm), supérieur à 20 cm (lớn hơn 20 cm) (greater than 20 cm), 5 à 10 cm (5đến 10 cm) (5 to 10 cm)	
Coloration du chaume (Màu sắc thân) (Colouration of the culm)	unie (đồng nhất) (uniform)	unie (đồng nhất) (uniform)	unie (đồng nhất) (uniform)	unie (đồng nhất) (uniform)

Fig. 2 – The automatic comparison of descriptions displays in a visual table the characters which are common or different between two or more descriptions. Here the comparison of a "vernacular" entity and the species *Dendrocalamus giganteus.*

2.5 Computer-aided identification for Indochine bamboos

All the information collected in the project is already digitalized in a structured format. The automatic HTML export of Xper^2 and Xper^2 online free access keys will be combined to offer a e-bamboo-flora.

3. Conclusion

The project "Indochine Bamboos" is still progressing. Three new species have been detected and will be published.

Presently, the Centre's collection has about 350 specimens from Vietnam, Laos and Cambodia. Few additional field trips are planned to complete the live collection in the Plant Conservation Centre in Phu An with typical bamboo species in Indochina.

The validation of the approach to identify vernacular names to scientific names could be proposed for other taxa, and made more automated.

Two master students and a PHD student are working on the subject. The project also offers the opportunity of organising training courses on the identification tools for students and young researchers coming from the participant institutions. Two workshop trainings for using the software Xper^2 were organized in 2008 and 2010. The participants were from Vietnam, Laos and Cambodia. This type of tools is attracting students interested in botany, enhancing their capabilities to analyse characters and taxonomic data.

Acknowledgement

The authors wish to thank Mrs Dransfield, Florian Causse and all participants of the "Bambous d'Indochine" project. This work is supported by the French initiatives Sud-Expert Plantes granted by the French. Ministry of Foreign Affairs.

References

[1] M. H. Diep, *et al.*, *Collection des variétés de bambou du Viet Nam.* Rapport scientifique après 3 ans de prospection des Bambousa du Viet Nam, 155 pp., 2005.

[2] M. H. Diep and M. L. Nguyen thi, *Ethnobotanique du bambou du Viet Nam*. Rapport scientifique de la Conférence scientifique de l'Université des Sciences Naturelles, novembre 2006.

[3] M. H. Diep, *Biodiversité du Bambou du Viet Nam.* Rapport scientifique de la Conférence scientifique de l'Université des Sciences Naturelles, novembre 2006.

[4] S. Dransfield, and E. A. Widjaja, *Plant Resources of South-East Asia. No 7 – Bamboos*. Backhuys Publishers, Leiden. 189 pp., 1995.

[5] J. Lebbe and R. Vignes, "Modelling taxonomic description for identification". In: P. Bridges, P. Jeffries, D. R. Morse and P. R. Scott (eds.), *Information Technology, Plant Pathology and Biodiversity*, pp. 37-46, 1998.

[6] H. Le Comte, *Flore Générale de l'Indochine*. Editeur Masson et Cie, 630 pp., 1912-1923.

[7] H. N. Nguyen, *Bamboos of Viet Nam.* Agriculture Editions, 199 pp., 2005.

[8] P. Hoang Ho, *Flore du Viet Nam.* Montréal Edition. 735 pp., 1992.

[9] A. Camus, E. G. Camus and H. Lecomte, *Flore générale de l'Indochine*, Masson et Cie, Paris, pp. 581-650, 1912-1923.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 213-216.
ISBN 978-88-8303-295-0. EUT, 2010.

Identification tools as part of Feedsafety research: the case of ragwort

Leo W.D. van Raamsdonk, Patrick Mulder, Michel Uiterwijk

Abstract — Ragwort (*Senecio jacobaea*) and related species of the genus *Senecio* are the main source of pyrrolizidine alkaloids. These plants grow in road verges, meadows and production fields and they show up in parties of roughage: grass and alfalfa. Monitoring can be carried out during the field production and harvesting stages. The final objective is to reject parties with a too high contamination level. Identification tools can support the decision to accept or refuse materials for the food production chain. A ragwort model has been developed for the mobile application Determinator. This identification model includes the relevant objects (species of the genus *Senecio*), and a range of so called confusing objects in order to minimise the chance of false positive identifications.

Index Terms — Determinator, diagnosis, identification, ragwort.

1 Introduction

Safe feed is one of the cornerstones of a healthy food production chain, and as important side-effect it supports the desired welfare of husbandry animals. In a lot of cases in the history of feed and food production emerging risks were initially detected by visual surveillance. Also in a majority of those cases visual inspection was replaced by more dedicated chemical detection methods.

Nevertheless, new risks still emerge and visual inspection is still at hand at the same moment that surveillance is needed. Recent examples are *Ambrosia* seeds in bird feeds, packaging materials in overdue materials, precatory bean (*Abrus precatorius*, in 2009 included in legislation) and ragwort in roughage and in salads for human consumption. The well known problem of animal byproducts is still primarily based on visual control. Identification tools are an essential support for these diagnosing problems.

L. W. D. van Raamsdonk and P. Mulder are with RIKILT – Institute of food safety, P.O. Box 230, 6700 AE Wageningen, the Netherlands. E-mail: Leo.vanRaamsdonk@wur.nl, Patrick.Mulder@wur.nl.
M. Uiterwijk is with Alterra, 6700 AE Wageningen, the Netherlands. E-mail: Michel.Uiterwijk@wur.nl.

2 The Ragwort problem

2.1 Background

Ragwort (*Senecio jacobaea*) is one of the sources of pyrrolizidine alkaloids and it can occur in parties of roughage: grass and alfalfa. Pyrrolizidine alkaloids are toxic for animals and a long term lethal effect is observed, especially in horses. Many other species of *Senecio* (e.g. common groundsel, *S. vulgaris*) and species of the family Boraginaceae (*Symphytum*, *Echium*) will produce pyrrolizidine alkaloids as well [1], [2].

Monitoring of the ragwort / pyrrolizidine alkaloids problem can be effected at several stages in the feed production chain. Visual inspection can be carried out during the field production and harvesting stages, when fresh materials are still present. Chemical LC-MS/MS analysis can be applied to trade parties of dried and processed roughage. In any situation visual screening can be followed by chemical confirmation.

Fig. 1 – Inflorescence of ragwort, *Senecio jacobaea*.

2.2 Strategy

Monitoring of production fields or road sides where *Senecio* species might occur is effective during pre-harvesting and harvesting times. On the spot identification and qualitative risk assessment can be achieved with a mobile knowledge system. Based on the prevalence of ragwort or other species it can be decided to use a harvested party for feed production or to ignore and destroy contaminated parties. In this way, identification tools can be useful for early

warning systems, so that costs in subsequent parts of the production chain can be avoided.

2.3 Knowledge system Determinator

Support of monitoring in those pre-harvesting stages is provided by the knowledge system Determinator. A datamodel is developed for this knowledge system including five *Senecio* species, and a series of 22 different yellow-flowering species, which can be confused with ragwort. Determinator can be used in laboratory (Windows XP, Vista) as well as in field situations (Windows mobile).

Determinator is a program package that assists the user in "determining" or identifying an object. A final conclusion is reached by entering answers to questions associated to the objects included in the dataset used. A match is calculated between the object as described by the user and each of the targets included in the chosen dataset. The process of identifying an object is supported by the possibilities to browse the included targets, and to compare two targets in every combination.

Every target in a fully developed datamodel is available with a description, with one or more images, and with one or more states for every feature (character). The descriptions and images are used to document the targets after the option Browse. The lists of feature states are being used to Compare two targets, and to Identify an object chosen by the user.

Determinator and the ragwort datamodel are freely available: www.determinator.wur.nl/UK/. Some screenshots are shown in Fig. 2.

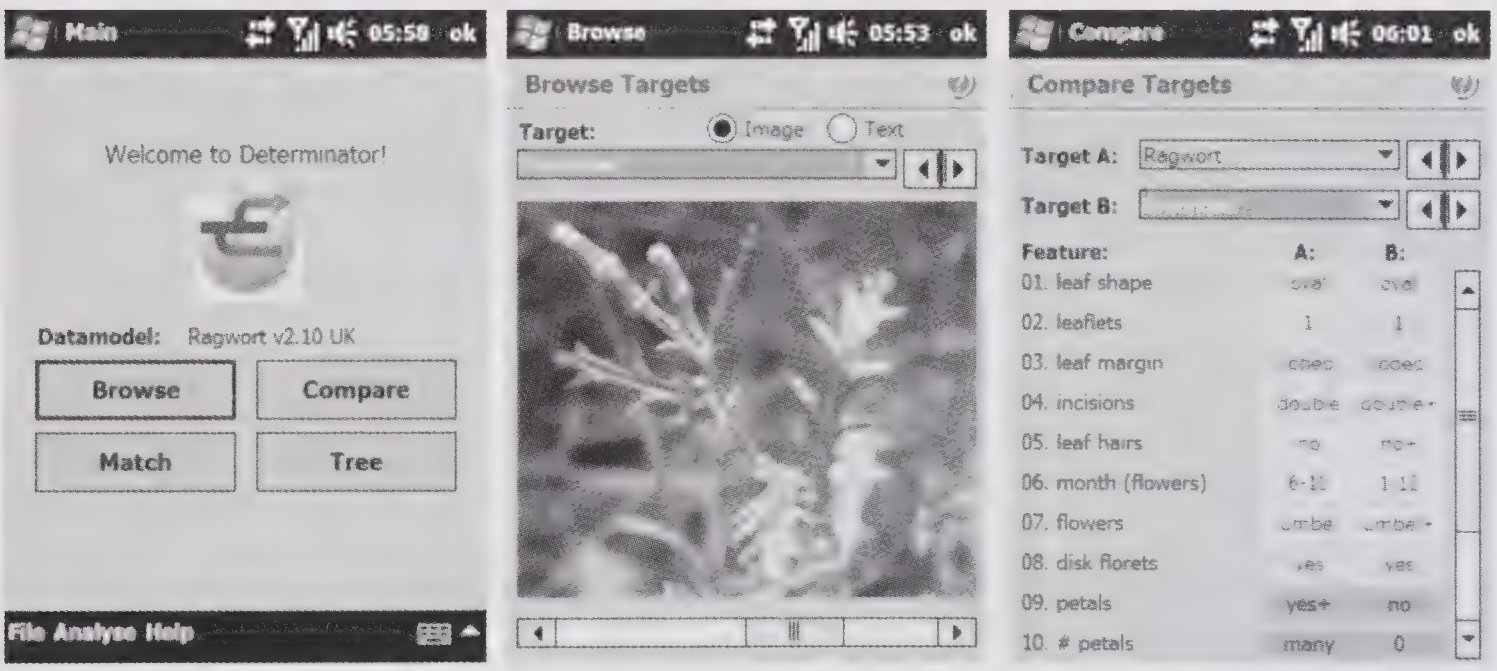


Fig. 2 – Screenshots of Determinator on a Windows mobile-based smartphone.

3 Discussion

The ragwort datamodel developed for Determinator is a highly dedicated identification tool. It is an example of an open classification model: only diversity is included that can directly support the final decision. Open classification models can function only in a situation where closed classification systems exist that

included all the existing diversity [3], [4], [5]. The flora of the Netherlands and the flora of the British Isles in Linneaus II [6] are examples of such classification systems that support the selected diversity in the ragwort datamodel. Another example of an open classification model is the decision support system ARIES [7] designed to support the ban on animal by-products as feeding stuff.

The philosophy of the ragwort datamodel and of ARIES is to include two types of objects. The first type of objects includes the species of *Senecio* or all types of animal by-products, respectively. The second type of objects added to the datamodel consists of a range of confusing objects. These confusing objects are meant to minimise false positive identifications.

Open classification models provide a good support for certain types of risk assessments, where information on identification is necessary. They can be developed in a relatively short time, exclusively targeted information should be included, and a connection exists with closed classification systems providing a full view on the relevant diversity.

References

[1] D. Frohne and H. J. Pfänder, *Poisonous Plants, second edition*, London, Manson Publishing, 2005.

[2] P. B. Pelser, H. de Vos, C. Theuring, K. Vrieling, T. Hartmann and T. I. Beuerle, "Frequent gain and loss of pyrrolizidine alkaloids in the evolution of *Senecio* section *Jacobaea* (Asteraceae)" *Phytochemistry*, vol. 66, pp. 1285–1295, 2005.

[3] L. W. D. van Raamsdonk, "The effect of domestication on plant evolution". *Acta. Bot. Neerl.*, vol. 44, pp. 421-438, December 1995.

[4] L. W. D. van Raamsdonk and T. de Vries, "Cultivar classification in *Tulipa*". *Acta Bot. Neerl.*, vol. 45(2), pp. 183-198, June 1996.

[5] W. L. A. Hetterscheid, R. G. van den Berg and W. A. Brandenburg, "An annotated history of the principles of cultivated plant classification". *Acta Bot. Neerl.*, vol. 45(2), pp. 123-134, June 1996.

[6] ETI bioinformatics. *Linnaeus II software package*. http://www.eti.uva.nl/, 2010.

[7] RIKILT Institute of food safety. *ARIES, Animal Remains Identification and Evaluation System*, http://aries.eti.uva.nl/, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 217.
ISBN 978-88-8303-295-0. EUT, 2010.

Two identification tools applied on Mascarene's corals genera (Xper²) and species (IKBS)

Yannick Geynet, Noël Conruyt, David Grosser,
Gérard Faure, David Caron

Abstract — For future biodiversity studies relying on species identification, environmental officers and researchers will only be left with monographic descriptions and collections in museums. This is why a knowledge base on the zooxanthellate scleractinian corals of the Mascarene Archipelago is being developed. This project offers results for both biologists/taxonomists and students or MPA-teams. Two online computer-based applications permit to identify genera and species. The first identification tool, called Xper², was developed by LIS (Informatic and Systematics Laboratory) in Paris, and is used for identifications to genera. The second tool, named IKBS (Iterative Knowledge Base System), was developed by IREMIA (Institute for Research in Applied Mathematics and Computer Science) in La Réunion, and is used for identifications from families to species. The tools presently work for *Astrocoeniidae*, *Pocilloporidae*, *Acroporidae* (only *Acropora* + *Isopora*), *Psammocoridae*, *Siderastreidae* (owns *Psammocoridae* as genera), *Fungiidae*, *Poritidae*, *Faviidae Faviinae*, *Faviidae Montastreinae*, *Mussidae*. We plan to start a new phase to add the last families, fully translate the web site in English and extend the Xper² identification to all the western Indian Ocean genera.

Index Terms — identification tool, IKBS, Mascarene archipelago, scleractinian corals, Xper².

Acknowledgement

This work was supported in most part by the EU (FEDER), the French Ministry of National Education, Advanced Instruction and Research, and the Réunion regional council. Web site: http://coraux.univ-reunion.fr/

Y. Geynet., N. Conruyt, D. Grosser and D. Caron are with the IREMIA lab from the Réunion University - PTU, 97490 Ste Clotilde – La Réunion. E-mail: yannick.geynet@univ-reunion.fr.
G. Faure is retired from the Univ. of Sciences Montpellier 2, 34000 Montpellier, E-mail: faure@cegetel.net.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 219.
ISBN 978-88-8303-295-0. EUT, 2010.

Interactive, illustrated, plant identification keys: an example for the Portuguese flora

Maria Helena Abreu Silva, Rosa Maria Ferreira Pinho, Lísia Graciete, Martins Pereira Lopes, Paulo Cardoso da Silveira

Abstract — At the University of Aveiro, a multimedia tool was developed to help the teaching of botany [1], [2]. It includes a Dichotomous Interactive Key (DIK), for 390 taxa of vascular plants occurring at "Ria de Aveiro" lagoon system. This key is linked to two different glossaries, one descriptive, the other illustrated. At the end of the identification process, the student is conducted to a webpage including photographs, descriptions and other relevant information about the taxa. The students considered it "very useful", and the successful identifications increased in the assessment tests in the two years after the introduction of the DIK. This seem to be an effective tool in the teaching of botany, namely in plant identification at secondary and post-secondary level.

Index Terms — interactive keys, illustrated keys, morphology, plant identification.

References

[1] P. Silveira, H. Silva, R. Pinho and L. Lopes, *Chaves Ilustradas. Identificação das plantas vasculares do Baixo Vouga Lagunar*, CD-ROM. Colecção Biorede, Universidade de Aveiro, ISBN: 972-789-211-6, 2006.

[2] H. Silva, R. Pinho, L. Lopes and P. Silveira, " Illustrated plant identification keys: an interactive tool to learn botany" *Computers & Education,* submitted for publication.

All authors are with the Department of Biology, University of Aveiro, 3810-193, Aveiro, Portugal E-mail psilveira@ua.pt., P. Silveira and H. Silva are also with the CESAM (Centre for Environment and Marine Studies), University of Aveiro, 3810-193, Aveiro, Portugal.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 221.
ISBN 978-88-8303-295-0. EUT, 2010.

The ORCHIS software used to identify 100 orchids species of Lao PDR

Pierre Bonnet, André Schuiteman, Boukhaykhone Svengsuksa, Daniel Barthélémy, Vichith Lamxay, Soulivanh Lanorsavanh, Khamfa Chanthavongsa, Pierre Grard

Abstract — The identification of plants, especially of orchids, is a major obstacle to the regulation of their trade. Lao PDR recently signed the CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora), and is facing many difficulties in its application at its borders, precisely because of lack of resources and information about its flora. This is particularly marked in Laos, where the pressures on the country's natural resources are particularly important because of the growth of this region. The National Herbarium Netherlands, the National University of Laos, and CIRAD have joined within the European project ORCHIS (http://www.orchisasia.org/) to develop, test and disseminate original tools to identify 100 orchids species in Laos. This identification tool, primarily targeted for managers of protected areas and customs authorities of the Country, is adapted to people with only a limited knowledge of botany. The tool, based on the IDAO software developed by CIRAD, enables the creation of a graphic sketch of the plant that is sought, thus allowing to overcome the constraints of language and knowledge of a specialized terminology.

Index Terms — orchids, Laos, biodiversity informatics, taxonomic database, identification tool.

P. Bonnet and D. Barthélémy are with INRA, UMR AMAP, Montpellier, F-34000, France. E-mail: Pierre.bonnet@cirad.fr, Daniel.barthelemy@cirad.fr.
A. Schuiteman (A.Schuiteman@kew.org) is with Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK.
B. Svengsuksa (svengpcy@laotel.com), V. Lamxay (vlamxay@yahoo.com), S. Lanorsavanh, and K. Chanthavongsa are with National University of Lao PDR, Faculty of Science of Lao PDR, Department of Biology, P.O. BOX 7322, Vientiane, Lao PDR.
P. Grard is with CIRAD, UMR AMAP, Montpellier, F-34000, France.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 223.
ISBN 978-88-8303-295-0. EUT, 2010.

A collaborative and distributed identification tool for plants

Philippe Laroche

Abstract — We are developing a website/smartphone application where users can deposit pictures of plants. Volunteers can then answer questions relating to these pictures: text questions, or drawing comparisons, or comparisons with other pictures. When sufficient data are available (50 or 100 of such answers) we can then propose an identification. The volunteers give answers based on "naive" observations of plant details. Example on a flower: do you see 3 petals, 4 petals, 5-8 petals, many petals ? For leaves, we display a small set of drawings, the set changes when the firsts questions provide sufficient information. To finish the identification, we allow for direct comparison with other pictures of similar detail. The questions are inside Iframe, and so are easily included on a network of participating web sites. Depending on the total number of visits to these sites, we shall have identifications within a few minutes or a few hours. We have now a set of 300 plants which are well-know in France. An important point about the method is that the volunteers help to classify the pictures we have as well as new pictures, so that the system is bootstrapping and scales easily. Moreover, we can easily experiment with new questions or drawings, to find those that reduce entropy most rapidly.

Index Terms — collaborative tool, PDA application, biodiversity collection building.

The author is with Agoralogie, 6 rue de Candie, F 75011 Paris, France. E-mail: philippe.laroche@agoralogie.fr.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 225-229.
ISBN 978-88-8303-295-0. EUT, 2010.

Alternative 2D and 3D Form Characterization Approaches to the Automated Identification of Biological Species

Norman MacLeod

Abstract — Few have sought to compare the performance of alternative types of morphological data for biological species identification. This investigation contrasts results of form characterization via form factors, superposed landmark coordinates, landmark-registered semilandmark outlines, 3D semilandmark networks, and raw digital images for a test set of seven Recent planktonic foraminifer species. While all data types performed better than the qualitative assessment of morphological variation by human taxonomists, landmark-registered semilandmark outlines and raw digital images delivered the best performance in the context of approaches that could reasonably serve as the basis for fully automated species identification systems.

Index Terms — automated identification, landmark coordinates, Foraminifers, taxonomy.

The automated identification of biological objects (individuals) and/or groups (e.g., species, guilds, characters) has been a dream of systematists' for centuries. The goal of some of the first multivariate biometric methods was to address the perennial problem of group identification and inter-group characterization [1], [2]. Despite much preliminary work in the 1950s and 60s, progress in designing and implementing practical systems for fully automated specimen identification has proven frustratingly slow. However, as recently as 2004 Dan Janzen updated the dream for a new audience [3].

> "The spaceship lands. He steps out. He points it around. It says 'friendly–unfriendly-edible–poisonous–safe–dangerous–living–inanimate'. On the next sweep it says '*Quercus oleoides–Homo sapiens–Spondias mombin–Solanum nigrum–Crotalus durissus–Morpho peleides–* serpentine'. This has been in my head since reading science fiction in ninth grade half a century ago." (p. 731)

The author is with the Palaeontology Department, The Natural History Museum, Cromwell Road, London SW7 5BD, N.MacLeod@nhm.ac.uk.

Janzen's solution to this classic problem involved building machines to identify species from their DNA. His predicted budget and proposed research team are "US$1 million and five bright people." (p. 731). However, recent developments in computer architectures, as well as innovations in software design, have placed the tools needed to realize Janzen's vision in the hands of the scientific community not in several years hence, but now; and not just for DNA barcodes, but for digital images of organisms.

A recent survey of small-scale automated species identification system trials (<50 taxa), shows an average reproducible accuracy of over 85 percent with no significant correlation between accuracy and the number of included taxa or the type of group being assessed (e.g., butterflies, moths, bees, pollen, spores, foraminifera, dinoflagellates, vertebrates) [4]. These figures should be compared with the disturbingly few blind test studies of accuracy and consistency of human taxonomist identifications that have been published to date [5], [10]. Human cognition studies [5] suggest that human experts who are routinely engaged in particular discriminations can return accuracies in the range of 84 to 95 percent. But in the (far more common) cases in which trained personnel must deliver identifications for species they are not dealing with on a day-to-day basis self-consistencies drop to 67-83 percent and consensus consistencies between identifiers to 43 percent. Moreover, semi-automated and automated identifications–often involving thousands of individual specimens–can be made in a fraction of the time required by human experts and can be done on site, on demand, anywhere in the world.

Is there a need for such systems? After all, biology has been getting by without them for millennia. What makes anyone think computers can – much less should – replace human taxonomists or that the taxonomic communities efforts would not be better spent lobbying for increased government funding for tried and true traditional α-taxonomy?

If evidence existed to reassure the scientific community that most taxonomic identifications are accurate and consistent current identification practices situation might be tolerable. There is little such evidence. For example, 1997 a group of geologists organized a blind test to try to resolve a controversy over whether marine animals went extinct before or after the meteorite impact that marked the end of the Cretaceous Period.[9] Four taxonomic experts were asked to identify species of microscopic foraminifera in a set of rock samples without being told the age of the samples. No consensus on when the animals died out was established – not because of any flaw in the test's design, but because the species lists produced were so different as to be incomparable, in some cases with just 25 percent of species names in common.

Contrary to some voices within the systematics community, these developments could not have come at a better time. As all scientists already know, the world is running out of specialists who can identify the very biodiversity whose preservation has become a global concern. In commenting on this problem in palaeontology as long ago as 1993, Roger Kaesler recognized[12] …

> "… we are running out of systematic paleontologists who have anything approaching synoptic knowledge of a major group of

> organisms" (p. 329). "Paleontologists of the next century are unlikely to have the luxury of dealing at length with taxonomic problems … [Paleontology] will have to sustain its level of excitement without the aid of systematists, who have contributed so much to its success." (p. 330).

This expertise deficiency cuts as deeply into those commercial industries that rely on accurate identifications (e.g., agriculture, biostratigraphy) as it does into a wide range of pure and applied research programmes (e.g., conservation, biological oceanography, climatology, ecology).

If truth be told, it is commonly, though informally, acknowledged that the technical, taxonomic literature of all organismal groups is littered with examples of inconsistent and incorrect identifications. This is due to a variety of factors, including taxonomists being insufficiently trained and skilled in making identifications (e.g., using different rules-of-thumb in recognizing the boundaries between similar groups), insufficiently detailed original group descriptions and/or illustrations, inadequate access to current monographs and well-curated collections and, of course, taxonomists having different opinions regarding group concepts. Peer review only weeds out the most obvious errors of commission or omission in this area, and then only when an author provides adequate representations (e.g., illustrations, recordings, gene sequences) of the specimens in question.

Few have sought to compare the performance of alternative types of morphological data for biological species identification. This investigation contrasts results of form characterization via form factors, superposed landmark coordinates, landmark-registered semilandmark outlines, 3D semilandmark networks, and raw digital images for a test set of seven Recent planktonic foraminifer species. While all data types performed better than the qualitative assessment of morphological variation by human taxonomists, landmark-registered semilandmark outlines and raw digital images delivered the best performance in the context of approaches that could reasonably serve as the basis for fully automated species identification systems.

Systematics too has much to gain, both practically and theoretically, from the further development and use of automated identification systems. It is now widely recognized that the days of systematics as a field populated by mildly eccentric individuals pursuing knowledge in splendid isolation from funding priorities and economic imperatives are rapidly drawing to a close. In order to attract both personnel and resources, systematics must transform itself into a "large, coordinated, international scientific enterprise" [13] (p. 4). Many have identified use of the internet–especially via the world-wide web–as the medium through which this transformation can be made. While establishment of a virtual, GenBank-like system for accessing morphological data, audio clips, video files and so forth would be a significant step in the right direction, improved access to observational information and/or text-based descriptions alone will not address either the taxonomic impediment or low identification consistency issues successfully. Instead, the inevitable subjectivity associated with making critical decisions on the basis of qualitative criteria must be reduced, or at the

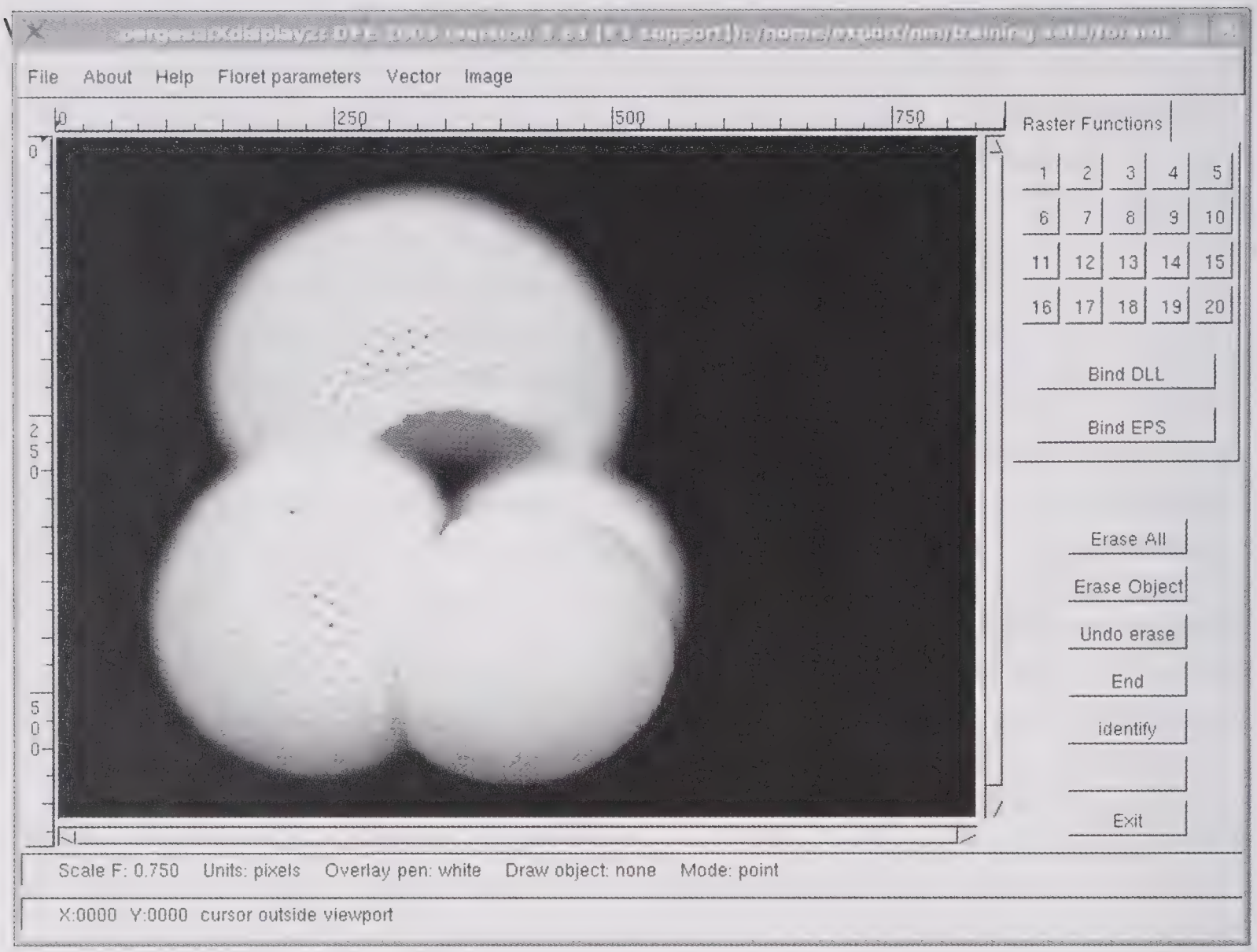


Fig. 1 – Example of the DAISY system interface displaying a planktonic foraminifer specimens from the test dataset. For this group, chamber arrangement, primary aperture position, and wall texture are among the primary taxonomic characteristics used to identify species.

Properly designed, flexible, and robust, automated identification systems, organized around distributed computing architectures and referenced to authoritatively identified collections of training set data (e.g., images, gene sequences) can, in principal, provide all systematists with access to the electronic data archives and the analytic tools necessary to handle routine identifications of common taxa. Properly designed systems can also recognize when their algorithms cannot make a reliable identification and refer that image to a specialist. Such systems will, inevitably, include elements of artificial intelligence that will allow them to improve their performance the more they are used. Most tantalizingly, once morphological (or molecular) models of a species have been developed and demonstrated to be accurate, these models can be queried to determine which aspects of the observed patterns of variation and variation limits are being used to achieve the identification, thus opening the way for the discovery of new and (potentially) more reliable taxonomic characters.

As has been demonstrated repeatedly through human history, scientific progress lies, in part, in constructing machines that do what machines do best and allowing humans to do what humans to best. Far from making taxonomists obsolete, the creation of automated identification systems will free them from the drudgery of delivering routine identifications to focus on the more conceptually

difficult issues of discovering, revising and describing species concepts, understanding how species fit into higher taxonomic and ecological groups and establishing how species function within natural systems. Getting high-throughput machine-learning systems on the agenda of research communities and scientific research funding councils, as well as into the study programmes of all sorts of disciplines, is required if taxonomy is to regain the sense of mission that will allow it to fulfil its potential as a twenty-first century science.

References

[1] R. R. Sokal and P. A. Sneath, *Principles of numerical taxonomy*. W. H. Freeman, San Francisco, 1963.

[2] P. H. A. Sneath and R. R. Sokal, *Numerical taxonomy: the principles and practice of numerical classification*. W. H. Freeman, San Francisco, 1973.

[3] Janzen, D. H., "Now is the time". *Philosophical Transactions of the Royal Society of London*, Series B, vol. 359, pp. 731–732, 2004.

[4] K. J. Gaston and M. A. O'Neill, "Automated species identification–why not?" *Philosophical Transactions of the Royal Society of London*, Series B, vol. 359, pp. 655–667, 2004.

[5] P. F. Culverhouse, R. Williams, B. Reguera, V. Herry and S. González-Gils, "Do experts make mistakes?" *Marine Ecology Progress Series,* vol. 247, pp. 17–25, 2003.

[6] W. P. Colquhoun, "The effect of a short rest pause on inspection efficiency". *Ergonomics,* vol. 2, 367–372, 1959.

[7] W. J. Zachariasse, W. R. Riedel, A. Sanfilippo, R. R. Schmidt, M. J. Brolsma, H. J. Schrader, R. Gersonde, M. M. Drooger and J. A. Brokeman, *Micropaleontological counting methods and techniques – an exercise on an eight meters section of the Lower Pliocene of Capo Rossello, Sicily*. Utrecht Micropaleontological Bulletins, vol. 17, pp. 1–265, 1978.

[8] R. Simpson, P. F. Culverhouse, R. Ellis and R. Williams. Classification of *Euceratium gran*. pp. 223-230. Neural Networks. IEEE International Conference on Neural Networks in Ocean Engineering. IEEE, Washington, D. C., 1991.

[9] R. N. Ginsburg, "Perspectives on the blind test". *Marine Micropaleontology,* 29, pp. 101–103.

[10] Kelly, M. G. 2001. "Use of similarity measures for quality control of benthic diatom samples". *Water Research,* vol. 35, pp. 2784–2788, 1997.

[11] K. W.Gobalet, "A critique of faunal analysis; inconsistency among experts in blind tests". J*ournal of Archaeological Science,* vol. 28(4), pp. 377-386, 2001.

[12] R. L. Kaesler, "A window of opportunity: peering into a new century of paleontology". *Journal of Paleontology,* vol. 67, pp. 329–333, 1993.

[13] Q. D. Wheeler, "Transforming taxonomy". *The Systematist,* vol. 22, pp. 3–5, 2003.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 231-236.
ISBN 978-88-8303-295-0. EUT, 2010.

VeSTIS: A Versatile Semi-Automatic Taxon Identification System from Digital Images

Nikos Nikolaou, Pantelis Sampaziotis, Marilena Aplikioti, Andreas Drakos, Ioannis Kirmitzoglou, Marina Argyrou, Nikos Papamarkos, Vasilis J. Promponas

Abstract — In this work we present a flexible Open Source software platform for training classifiers capable of identifying the taxonomy of a specimen from digital images. We demonstrate the performance of our system in a pilot study, building a feed-forward artificial neural network to effectively classify five different species of marine annelid worms of the class Polychaeta. We also discuss on the extensibility of the system, and its potential uses either as a research tool or in assisting routine taxon identification procedures.

Index Terms — digital image analysis, open source, semi-automatic taxon identification.

1 Introduction

Automated taxon identification (ATI) can be defined as the process of automating the routine identification of specimens [1] through the exploitation of modern computer science technologies and domain knowledge. ATI methods are based on mathematical descriptors of morphological [1], [2], [3], behavioural [4] or genetic [5] characters. These data are used as input into pre-processing and analysis pipelines, which are most often based on statistical or machine learning methods. ATI procedures are quickly becoming a necessity in the effort to understand and monitor global biodiversity.

So far, several research efforts to deal with ATI from digital images have been

N. Nikolaou, I. Kirmitzoglou, V.J. Promponas are with the Bioinformatics Research Laboratory, Department of Biological Sciences, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus. E-mail: nkleopas@gmail.com, bip6ki2@ucy.ac.cy, vprobon@ucy.ac.cy.
M. Aplikioti, A. Drakos, M. Argyrou are with the Department of Fisheries and Marine Research, 101 Vithleem Street, 1416 Nicosia, Cyprus. E-mail: maplikioti@dfmr.moa.gov.cy, andreas_drakos@hotmail.com, margyrou@dfmr.moa.gov.cy.
P. Sampaziotis, N. Papamarkos are with the Department of Electrical and Computer Engineering, Democritus University of Thrace, 67100 Xanthi, Greece. E-mail: psampaz@gmail.com, papamark@ee.duth.gr.

reported [2], with three major ones focusing on the implementation of semi-automatic species identification systems; (i) DAISY [3], (ii) SPIDA web [6], and (iii) ABIS [7]. Important drawbacks of such systems are that they are either suitable for a relatively narrow taxonomic range (e.g., SPIDA, ABIS) or unavailable for public use (e.g. DAISY, ABIS). Nevertheless, both these shortcomings could be eliminated in a community-based approach with the availability of suitable extensible platforms open for further development. Extensibility can be achieved in a dual manner: (i) at the software component level (e.g. by an Open Source modular software), and (ii) at the data level, with a flexible scheme to permit incorporation of novel data types regarding the taxonomic range accepted by the system, or data and feature types utilized in the ATI task.

In this work, we present our progress in designing and implementing such an Open Source computer system, VeSTIS. We demonstrate VeSTIS in the systematic identification of 5 species of the Class Polychaeta (Phylum Annelida), a marine macroinvertebrate group well known for the identification difficulties it presents.

2 Materials and Methods

2.1 Description and Key Features of the System

VeSTIS is intended to be a generic user-friendly platform capable of virtually identifying any taxonomic unit. It currently embeds a large number of state-of-the-art digital image analysis, enhancement and pattern recognition algorithms making it independent from the use of commercial software. Moreover, VeSTIS incorporates an SQL-based database schema and client-server technology to allow multiple users working simultaneously. The database schema was specifically designed to aid easy storage and retrieval of meta-data and to allow publishing of its contents on the Internet. This adds to the extensibility of the platform by facilitating the development of web-modules, such as a national biodiversity portal or a web-application for remotely identifying specimens through the users' browser. Finally, a very important characteristic is the ability to train VeSTIS with user-selected features in order to optimize the ATI process.

2.2 Species Selection, Sample Collection and Image Acquisition

In order to test the functionality of the system, five Polychaete species were used: *Nematonereis unicornis* (Smarda, 1861), *Marphysa bellii* (Audouin & Milne-Edwards, 1833), *Polyophthalmus pictus* (Dujardin, 1839), *Armandia polyophthalma* (Kükenthal, 1887) and *Terebellides stroemi* (Sars, 1835). These species were selected due to: (i) their high abundance in the coastal waters of Cyprus, and (ii) the relatively few problems in their identification compared to other Polychaete species.

Samples were collected with a Van Veen grab from a number of coastal sampling stations at depths of 25-35m in soft substrates. They were then sieved with a 0.5mm sieve, fixed and properly preserved. Finally, all Polychaete

specimens were identified to species level with the use of stereoscopes and microscopes.

Prior to finalising the exact photo-shooting conditions, we evaluated a series of factors directly related to the quality of the shots; i.e. various magnifications, background colour, lighting source and homogeneity, specimen body parts and their orientation, as well as specimen fixation. The best results were obtained by fixing the specimens between slides against a uniformly black background. For illuminating the system we used two Leica CLS150X cold light sources with the optic fibres oriented in a way that minimized shadows. For this demonstration, we focused on the frontal body part of the animals and specifically on the head and the first 10 segments.

All images used for training and validating VeSTIS were acquired using a Leica DFC290 camera mounted on a Leica MZ7.5 stereo-microscope. Photos were taken under specimen-size dependent magnifications (in the 12.6x-32x range) with the maximum resolution supported by the camera (3.2 MP) through the Leica Application Suite (LAS) software. Image pre-processing was carried out within VeSTIS.

2.3 Image Pre-processing and Feature Extraction

***Object (specimen) orientation correction*:** Image orientation is corrected, for the specimen to lay in a horizontal direction (Fig. 1A and 1B). This is important for object contour representation (see below).

***Image segmentation and object isolation*:** In order to isolate the object in the image, we used the Otsu binarization method [8]. This is a segmentation process which automatically creates a black (object) and white (background) image (Fig. 1C) based on the image histogram. Using connected component analysis, VeSTIS locates and isolates the object.

Object contour representation & feature vector generation: Upper/lower object profile features are computed by recording the distance of the lower boundary of the bounding box to the furthest/closest object pixel for each image column.

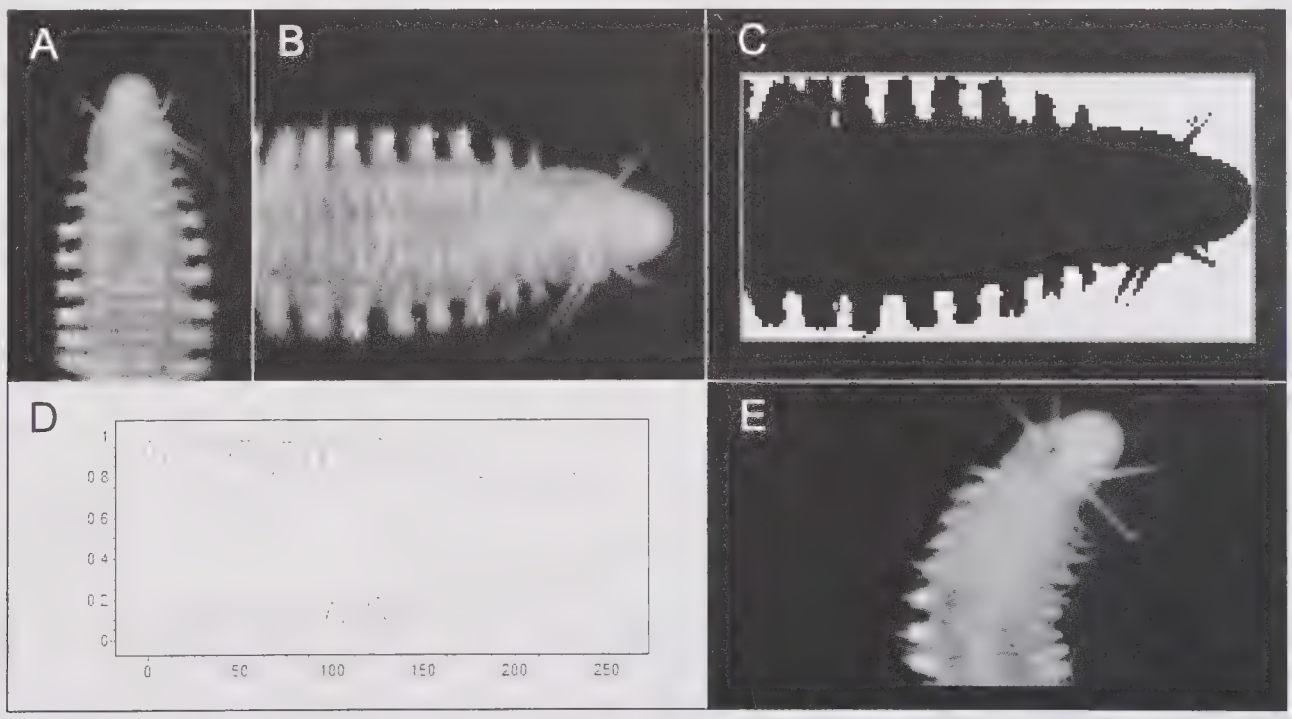


Fig. 1 – (A) Original, and (B) Corrected specimen orientation. (C) Segmentation and object isolation. (D) Contour representation. (E) An image classified as *bad* due to curvature.

All values vary between 0 and 1, since they are normalized by the height of the object (Fig. 1D). These two profiles can be considered to form a closed curve, allowing the use of Fourier descriptors [9] to mathematically describe the object's contour. Fourier descriptors allow bringing the power of Fourier theory to shape parameterisation by characterising a contour with a set of numbers that represent the frequency content of a whole shape. They are invariant to rotation, scale, and translation and are used as the input vector for the feed-forward artificial neural network (FFANN).

2.4 Generation of training and validation data sets

For generating training/validation sets, we manually classified all images based on species, specimen, orientation and condition. For four of the species in question, orientation was either dorsal or ventral. For *T. stroemi* only lateral-view photos were taken, mainly because of the species' morphology. Images were classified as *good* (G) whenever the specimen was in a good condition or *bad* (B) if the specimen was curved or moderately destroyed (Fig. 1E). We then created 3 training sets based on specimens' orientation, using only images flagged as *good*. Following a similar procedure we generated 9 different sets for evaluation purposes using both *good* & *bad* images (Tab. 1). We only included *good* images in training sets to reduce noise and test the ability of our approach to correctly classify problematic images/specimens. Multiple images were acquired for each specimen. However, a single image of each individual was included in either the training or the validation sets, in order to (i) avoid over-fitting during training, and (ii) minimize any bias on the estimation of the performance of the classifier. Thus, any pair of training-validation sets was strictly disjoint.

Training sets			Validation sets								
TS1	**TS2**	**TS3**	**VS1**	**VS2**	**VS3**	**VS4**	**VS5**	**VS6**	**VS7**	**VS8**	**VS9**
DLG	VLG	DVLG	DLG	VLG	DVLG	DLB	VLB	DVLB	DLGB	VLGB	DVLGB

Tab. 1 – Image types included in different training and validation data sets. D, V, L = Dorsal, Ventral, Lateral view; G, B = Good, Bad image classes.

3 Results and Discussion

Two FFANNs were trained for each training set in batch mode with the resilient back-propagation learning algorithm [10], each initialized with different random weights. A fully connected architecture, with a single hidden layer of 30 neurons and a sigmoid activation function, proved to be good choices after experimentation. Five output units served for classifying each specimen to the respective species using a 'winner-take-all' output encoding scheme. FFANNs were trained for 2000 epochs, and in all cases the mean squared error of desired versus predicted outcomes of the networks converged to very small values. The

performance of each FFANN was evaluated with the independent validation sets (Tab. 2). We also observed the performance of a simple ensemble average of independent classifiers trained with different types of data. In several cases the performance was drastically improved (Tab. 2).

	FFANN-TS1	FFANN-TS2	FFANN-TS3	ENSEMBLE
VS1	**0.702 (0.017)**	0.667 (0.034)	0.702 (0.051)	0.738
VS2	0.693 (0.016)	**0.727 (0.000)**	0.727 (0.000)	0.705
VS3	0.698 (0.000)	0.698 (0.016)	**0.715 (0.025)**	0.709

Tab. 2 – Evaluation of identical FFANNs trained with different training sets on independent validation data sets. For each FFANN the performance reported corresponds to the average overall performance (and standard deviation) of two independently trained networks initialized with different random weights. A simple ensemble averaging approach often seems to outperform individual classifiers. Data sets are described in Tab. 1. Specimen orientation seems to be an important factor affecting classification accuracy. As expected, results obtained with *bad* images were clearly inferior (data not shown).

Species identification is a painstaking and time-consuming task, which requires highly skilled and adequately trained scientific personnel. Although the design and implementation of reliable and accurate ATI methods is a challenging problem, it will definitely give rise to more experimentation and thus to the growth and evolution of systematics. It is anticipated that Open Source solutions will boost development, applicability and usage of ATI methods similar to what has been experienced in the field of computational molecular biology.

We are currently working on adding more software components to VeSTIS (feature extractors, classifiers, etc.). We specifically plan to address the feature selection task, since classification quality is expected to depend mainly on the features rather than the classifier. This is also an attempt to cover a gap in the literature that mainly deals with the effectiveness of classifiers.

VeSTIS, although currently in alpha phase, is being actively developed. We expect to release the first beta binaries and source code at the url http://troodos.biol.ucy.ac.cy/BRL/ within late 2010.

Acknowledgement

This work was co-funded by the Republic of Cyprus and the EU European Regional Development Fund (ERDF) through a grant from the Cyprus Research Promotion Foundation (AEIFORIA/FISI/0308(BIE)/10).

References

[1] K. J. Gaston and M. A. O'Neill, "Automated Species Identification: Why Not?", *Philos. Trans. R. Soc. Lond. B. Biol. Sci.,* vol. 359, pp. 655-67, 2004.

[2] N. MacLeod, *Automated Taxon Identification in Systematics: Theory, Approaches and Applications*. Boca Raton, FL: CRC Press, 2008.

[3] A. T. Watson, M. A. O'Neill and I. J. Kitching, "Automated Identification of Live Moths

(Macrolepidoptera) Using Digital Automated Identification System (Daisy)", *Systematics and Biodiversity,* vol. 1, pp. 287-300, 2004.

[4] J. Tanttu, J. Turunen, A. Selin and M. Ojanen, "Automatic Feature Extraction and Classification of Crossbill (*Loxia* spp.) Flight Calls", *Bioacoustics,* vol. 15, p. 251, 2006.

[5] A. Valentini, F. Pompanon and P. Taberlet, "DNA Barcoding for Ecologists", *Trends in Ecology & Evolution,* vol. 24, pp. 110-117, 2009.

[6] K. N. Russell, M. T. Do, J. C. Huff and N. I. Platnick, "Introducing Spida-Web: Wavelets, Neural Networks and Internet Accessibility in an Image-Based Automated Identification System". In: N. MacLeod (ed.), *Automated Taxon Identification in Systematics: Theory, Approaches and Applications*, Boca Raton, FL: CRC Press, pp. 131-152, 2008.

[7] T. Arbuckle, S. Schroder, V. Steinhage and D. Wittmann, "Biodiversity Informatics in Action: Identification and Monitoring of Bee Species Using Abis". In: *15th International Symposium Informatics for Environmental Protection*, Zurich, pp. 425-430, 2001.

[8] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", *IEEE Transactions on Systems, Man and Cybernetics,* vol. 9, pp. 62-66, 1979.

[9] O. Petkovic and J. Krapac, *Shape Description with Fourier Descriptors*, Technical Report, 2002.

[10] M. Riedmiller and H. Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The Rprop Algorithm". In: *IEEE International Conference on Neural Networks,* San Francisco, 1993.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 237-242.
ISBN 978-88-8303-295-0. EUT, 2010.

Iterative Search with Local Visual Features for Computer Assisted Plant Identification

Wajih Ouertani, Pierre Bonnet, Michel Crucianu,
Nozha Boujemaa, Daniel Barthélémy

Abstract — To support computer assisted plant species identification in a realistic, uncontrolled picture-taking condition, we put forward an approach relying on local image features. It combines query by example and relevance feedback to support both the localization of potentially interesting image regions and the classification of these regions as representing or not the target species. We show that this approach is successful, and makes prior segmentation unnecessary.

Index Terms — assisted identification, biodiversity informatics, local features, local query, object localization, relevance feedback.

1 Introduction

Given the large volume and increasing accessibility of biodiversity data - e.g. Encyclopaedia of Life [1], Atlas of living Australia [2], or ZipcodeZoo - gathered from all over the world, it is even more important to explore, master and capitalize this type of knowledge [3]. Joint efforts of biologists, information science and data-mining communities are required for solving significant common problems. As biological image databases are increasing rapidly [4], automated species identification based on digital data is of great interest for accelerating biodiversity assessment, research and monitoring [5]. We put forward here an interactive identification approach in which a botanist having a partially annotated a large image database is assisted by a Relevance Feedback search mechanism to identify a plant species. The botanist can then easily select the relevant unlabeled images (without having to go through the entire database) and label them at once with the name of the species.

W.Ouertani, M. Crucianu and N. Boujemaa are with INRIA, IMEDIA Project, BP 105, 78153 Le Chesnay cedex, France. E-mail: (Wajih.Ouertani, Michell.Crucianu, Nozha.Boujemaa)@ inria.fr.
P. Bonnet, D. Barthélémy and W.Ouertani are with INRA, Amap Joint research unit, CIRAD,TA A-51/PS2, 34398 Montpellier cedex 5, France. E-mail: (Pierre.Bonnet, Daniel.Barthelemy)@cirad.fr.

2 Context and resulting challenges

2.1 Content-based image retrieval and interactive identification

In a query by visual example (QBVE), an example image is first provided to the search engine as a visual query. The engine returns images that are visually similar to the query image, using a metric on the space of the low-level features that represent the images. Motivated by the "semantic gap" issue, *i.e.* the fact that such features seldom reflect the user's intention, a Relevance Feedback (RF) [6] mechanism includes the user in the retrieval process. In a RF session, the search result is iteratively refined. For a given query, the system first retrieves a set of images ranked according to the predefined similarity measure between the query vector and feature vectors of images in the database. Then, the user provides feedback regarding this result, by qualifying the returned images as either "relevant" or "irrelevant". From this feedback, the engine iteratively learns the visual features of the images, and returns improved results to the user. A good RF mechanism should find the user intention with minimal interaction [7].

This retrieval refinement technique was applied to botanical databases with pictures taken in controlled conditions [8], but it has important limitations resulting from the global image description. To remove such restrictions on picture-taking conditions, we extend here RF to the use of *local* features (LF). This is a more adequate representation of image regions, which allows users to provide a precise feedback by freely selecting relevant and irrelevant *regions of interest* in images.

2.2 Challenges

We address here learning and recognition challenges that come from strong variations in viewpoint, picture-taking conditions, interactivity and generalization requirements. Recent work on plant species identification requires reliable prior segmentation of informative organs such as leaves [9], [10] (with controlled picture-taking conditions) or flowers [11] (less restrictive conditions). With such well-controlled pictures, the shape of a leaf, its margins, or several local and region-based features of flowers are employed for recognition. In general, due to variations in the natural environment, plant accessibility, picture-taking system and intention, an object of interest (a plant or a plant part) may appear on different backgrounds and cover a potentially small part of the image (see first row in Fig. 1). This supports the use of LF to focus on the target object. Also, in a botanical identification context, some images illustrate global aspects of a plant or of an inflorescence, while others show details having different visual attributes. A same object of interest could thus be represented in various poses and at different scales (see second row in Fig. 1).

Relevance feedback brings in two additional challenges. First, the search engine should respect the interactivity requirement, *i.e.* quickly respond during each round. Even if joint object segmentation and recognition (*e.g.* [12]) could improve identification, its additional cost makes it inappropriate for interactive

retrieval. Second, at each RF round the user only labels *a few* images. For the retrieval session to be successful, the system should generalize well from these few examples. In the next section we propose an approach that addresses these problems using LF.

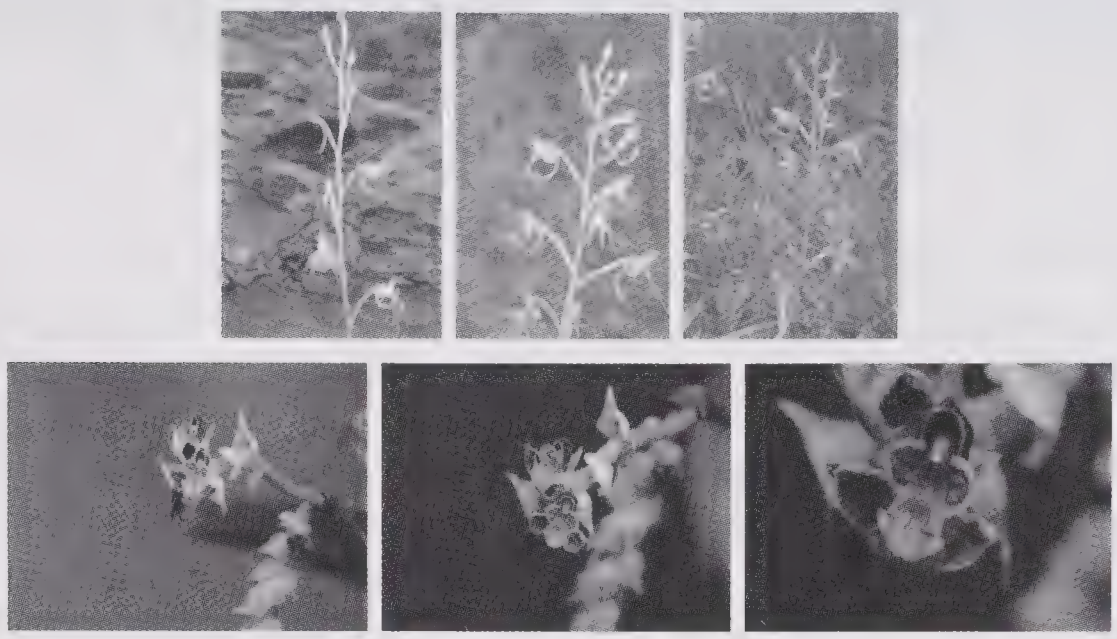


Fig. 1 – Background variations of an inflorescence of *Habenaria* species (1st row), scale and pose variation of an inflorescence of *Cleisomeria lanatum* (Lindl. ex G.Don, 2nd row).

3 Identification Approach

We propose to jointly use search by example with local queries and supervised classification (with Support Vector Machines, SVM). Every RF round thus consists of two stages: (1) QBVE using as query the LF that were previously found relevant; (2) result re-ranking by the SVM decision function, applied to the potentially relevant set of features in every returned image. This joint use of QBVE and SVM classification serves two purposes. First, it allows to locate, in the returned images, the potential regions of interest (see Fig. 2, green and red points) that have to be evaluated by the SVM. A region of interest is here the set of LF that were found to be individually similar to some LF in the query. An image can indeed contain objects from multiple classes; our approach will focus on the potentially relevant parts and ignore other, irrelevant parts (blue points in Fig. 2). In this context, the task of the SVM is to solve ambiguity and distinguish sets of LF that belong to the target specie (Fig. 2, middle) from sets composed of LF that are individually similar to relevant LF but, when considered together, do not correspond to the target species (Fig. 2, right).

Second, QBVE can be very fast with an appropriate index structure - we rely here on *a posteriori* multi-probe locality sensitive hashing [13] - and only images containing hit points (*i.e.* points that are individually similar to relevant LF) have to be evaluated by RF rather than all the images in the database, which significantly improves scalability.

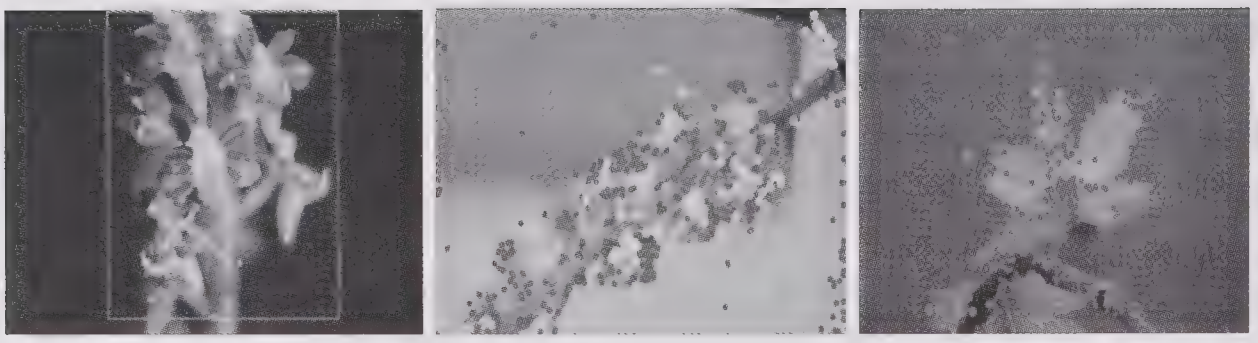


Fig. 2 – Region of interest localization: user target (left) and two candidate images with LF belonging to the target (middle) or not (right). The other LF are ignored.

We assume that the distribution of LF in the selected sets brings relevant discriminating information with respect to the joint presence of LF, so we employ the pyramidal matching kernel (PMK, [14]) or the kernel based on random histograms (RH, [15]). The SVM has thus to downgrade image regions (sets of LF) whose LF are individually similar to LF of the target species, but whose distribution does not correspond to this target.

4 Experimental Evaluation

We employed two different image databases for the evaluations. The first one was produced by AMAP Joint Unit on Laos orchid's reproductive organs (mainly inflorescences and flowers). It contains 1913 images for 181 orchid species. There are significant variations in scale, pose and lighting (see Fig. 1, 2). Botanists manually labelled 2347 regions of interest. The second database is Oxford flowers 17 (www.robots.ox.ac.uk/~vgg/data/flowers/17/), consisting of 17 flower categories with 80 images each. The database includes common UK flowers; there is a significant variation within a same class and close similarity between several classes. There is also a ground truth showing fine flower segmentation for a subset of the images [11].

We compare RF with global image description (`GF_RF`) to RF with local descriptions (`LF_RF_QVE_Harris`, `LF_RF_QVE_SIFT`). The global image description employed (named "joint description" below) concatenates a Laplacian weighted RGB histogram, a Fourier-based histogram and a Hough histogram [2]. Two types of LF were employed: (i) joint description (with coarser histograms) obtained in the neighbourhood of Harris colour points, and (ii) SIFT [16]. The experiments were performed by using the ground truth to emulate user feedback under realistic conditions. Each RF session consists of 8 iterations. At every iteration, the emulated user labels the first 3 relevant and the first 3 irrelevant unlabelled regions. Fig. 3 shows the mean average precision (MAP) of system's responses where recall equals precision (MAP at R=P), for the three RF mechanisms. Only the 10 orchid classes having enough image examples were used for generating RF sessions. Fig. 3 (left) shows that, even with few iterations (1^{st} to 4^{th}, less than 50% of the available training data), RF with LF outperforms global RF. We also note that the results obtained with SIFT (features ignoring colour!) are better than those with Harris points whose description includes colour. This is due to the fact that scale and shape variations within a same class are more important than colour differences between classes in this dataset.

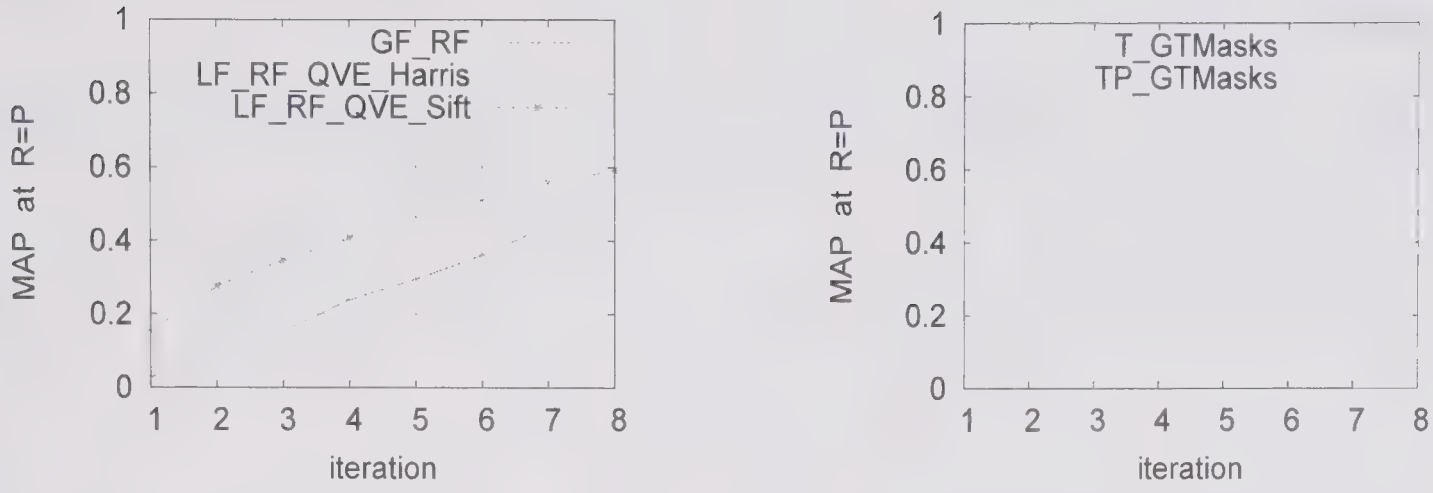


Fig. 3 – MAP evolution over RF iterations. Left: on Orchids database. Right: on Oxford flowers 17 database, with and without segmentation masks in prediction stage.

Using `LF_RF_QVE_Harris` and the fine segmentation ground truth provided in Oxford flowers 17 database, we performed two experiments in which we use segmented objects as training examples and, for the prediction stage, we either (i) use only hit points (retrieved by QBVE) that fall in pre-segmented objects of interest in a candidate image (`TP_GTMasks`), or (ii) use all the hit points retrieved in a candidate image (`T_GTMasks`). As can be seen in Fig. 3 (right), the object localisation given by the QBVE stage allows to reach a performance that is close to the one obtained with fine prior segmentation. We also find that the inclusion of a small part of object's neighborhood provides a relevant context that increases recognition accuracy.

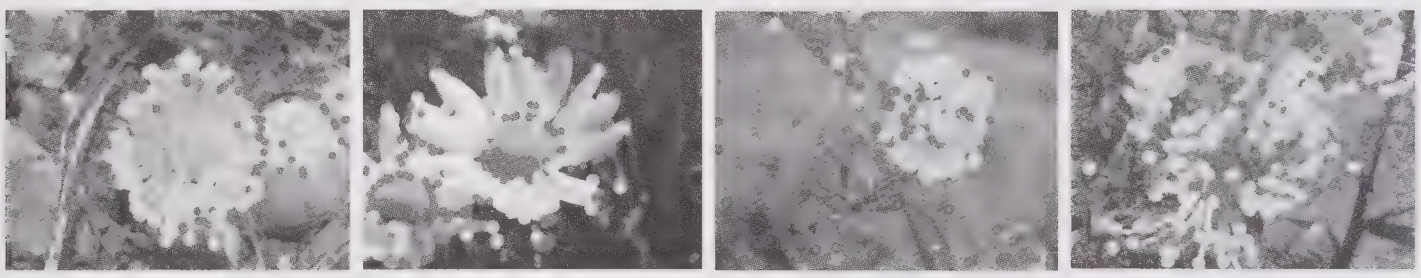


Fig. 4 – Object localization examples on Oxford Flower 17 database, points showing the object of interest. From left to right: Colt's Foot, Daisy Flower, Buttercup, Tiger Lily.

5 Conclusion

Content-based image search can provide a significant contribution to plant species identification. However, to make it successfully applicable to realistic contexts, we argue that it is necessary to let the user interact with the system on the basis of local image descriptions that allow to focus on the relevant part of an image. We proposed a relevance feedback method relying on local images features. It also makes use of an LF retrieval stage in order to locate potentially interesting image regions and improve scalability to larger image databases. We have shown that this approach can be successful and that it makes prior segmentation unnecessary. The results also show how important it is to devise local features that are robust to most of the variations that can be expected when pictures are taken in more general, uncontrolled conditions.

Acknowledgements

This work is part of the flagship project of Agropolis fondation: Pl@ntNet, http://www.plantnet-project.org.

References

[1] E. O. Wilson, "The encyclopedia of life", *Trends in Ecology and Evolution,* 18 (2), 2003.

[2] Anon., 2008, "Atlas of Living Australia – sharing biodiversity knowledge to shape our future", *Proc. R. Soc. Western Australia*, Nov. 2008.

[3] N. F. Johnson, "Biodiversity informatics". *Annu. Rev. Entomol.,* vol. 52, pp. 421-438, 2007.

[4] S. J. Baskauf and B. K. Kirchoff, "Digital plant images as specimens: toward standards for photographing living plants". *Vulpia*, vol. 7, pp. 16–30, 2008.

[5] K. J. Gaston and M. A. O'Neil, "Automated species identification: why not?" *Phil. Trans. R. Soc. B.*, vol. 359, pp. 655-667, 2004.

[6] X. S. Zhou and T. S. Huang, "Relevance feedback for image retrieval: a comprehensive review", *Multimedia Systems*, vol. 8, no. 6, pp. 536-544, 2003.

[7] M. Ferecatu, *Image retrieval with active relevance feedback using both visual and keyword-based descriptors*. PhD thesis, Université de Versailles, France, 2005.

[8] M. Coutaud, P. Bonnet, A. Joly, R. Enficiaud, N. Boujemaa and D. Barthélémy, "Advances in taxonomic identification by image recognition with the generic content-based image retrieval IKONA". In: *e-Biosphere 09: Intl. Conf. on Biodiversity Informatics*, London, 2009.

[9] I. Yahiaoui, N. Hervé and N. Boujemaa, "Shape-based image retrieval in botanical collections". In: *7th Pacific Rim Conf. on Multimedia*, LNCS vol. 4261, pp. 357–364, 2006.

[10] P. N. Belhumeur, D. Chen, S. Feiner, D. W. Jacobs, W. J. Kress, H. Ling, I. Lopez, R. Ramamoorthi, S. Sheorey, S. White and L. Zhang, "Searching the world's herbaria: A system for visual identification of plant species". In: *European Conf. on Computer Vision, LNCS vol.* 5305: pp. 116–129. Springer, 2008.

[11] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes". In: *6th Indian Conf. on Computer Vision, Graphics & Image Proc.*, pp. 722–729, Washington, DC, USA.IEEE Computer Society, 2008.

[12] J. Shotton, J. Winn, C. Rother and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context". *Int. J. Comput. Vision*, vol. 81(1), pp. 2–23, 2009.

[13] A. Joly and O. Buisson, "A posteriori multi-probe locality sensitive hashing". In: *16th ACM intl. conf. on Multimedia*, pp. 209–218, New York, NY, USA, 2008.

[14] K. Grauman and T. Darrell, "The pyramid match kernel: Efficient learning with sets of features". *J. Mach. Learn. Res.*, vol. 8, pp.725–760, 2007.

[15] W. Dong, Z. Wang, M. Charikar and K. Li, "Efficiently matching sets of features with random histograms". In: *16th ACM intl. conf. on Multimedia*, pp. 179–188, New York, NY, USA, 2008.

[16] D. G. Lowe "Distinctive image features from scale-invariant keypoints". *Intl. J. Comp. Vis.,* vol. 60(2), pp. 91-110, 2004.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 243-248.
ISBN 978-88-8303-295-0. EUT, 2010.

Image data banks and geometric morphometrics

Anna Loy, Dennis E. Slice

Abstract — This paper examines the opportunities offered by recent advances in digital image processing to allow access to natural history museum collections without direct handling of specimens. It specifically refers to two- and three-dimensional data recording and analysis in the frame of geometric morphometrics.

Index Terms — geometric morphometrics, 2D and 3D image recording, landmarks, osteological collections.

1 Introduction

Osteological collections, especially skull collections, represent an ideal material for the study of morphological variation in both time and space in a variety of vertebrates. Its multiple functional properties (protection of the brain and sense organs, feeding and respiratory structures) make the skull a highly informative structure where both highly conservative and plastic characters coexist [1]. These qualities led to a rich production on the intra- and interspecific variation of vertebrates based on skull features [1] and to a precise coding of traditional quantitative characters [2]. The geometric morphometrics *revolution* in the '90s [3], [4] offered a new powerful tool to investigate the variation of biological forms, allowing the distinction between size and shape variation. Geometric Morphometric (GM hereafter) studies have found their elective applications in the analysis of osteological collections devoted to clarification of clarify phylogenetic and evolutionary patterns among vertebrates, [5], [6], [7], [8]. Basic data for GM are usually recorded either from 2D or 3D images. Taking advantage of the tremendous advances in digital technologies, museums can play a fundamental role for future GM studies by offering an easy and rapid remote access to their collections [9], [10].

A.Loy is with the Università del Molise, Pesche, Italy, I-86090. E-mail: a.loy@unimol.it.
D.E.Slice is with the Florida State University, Tallahassee, FL 32306-4120. E-mail: dslice@fsu.edu.

2 Geometric Morphometrics

GM uses sets of Cartesian coordinates, such as (semi-)landmark locations, outlines, curves, and surfaces, to capture the geometric information about biological structures and preserves that information throughout the analyses, including the multivariate treatments of data [4], [5], [11]. Most multivariate methods of GM are linearizations of statistical analyses of distances and directions in the Kendall's shape space. Each point in this shape space represents the shape of a configuration of points (landmarks) in some Euclidean space, irrespective of size, position, and orientation [5]. In shape space, scatters of points correspond to scatters of entire landmark configurations (specimens), not merely scatters of single landmarks, and differences among shape configurations are most often expressed as cord distances relative to a curved, generalized Procrustes space [4], [12], [13], [14], [15].

Configurations are described by either two (x,y) or three (x,y,z) Cartesian coordinates of homologous points (landmarks). The advantage of working with 2D landmarks is that these data are easily recorded from digital pictures through easily accessible and friendly software, e.g., TpsDig [16]. Meanwhile, studies using coordinates of 3D points are becoming standard in some fields, such as physical anthropology [17], [18]. A distinct advantage of the use of 3D coordinates is that the definitions of landmark points are often much less arbitrary in three dimensions than they are in 2D projections [5]. An historical disadvantage of three dimensional landmarks is that they can only be recorded directly from the objects by means of devices like the 3D Microscribe or Polhemus digitizers or gathered from 3D pictures obtained from very expensive scanners. Unfortunately, statistical methods for dealing with such additional data types (surfaces and volumes) are still in their infancy. Moreover, we also still lack effective methods for the visualization of genuinely 3D shape variation whether for points or more complicated data structures [5].

3 2D and 3D digital images for GM analyses

Datasets used for GM analyses can be derived from 2D and 3D digital images (Fig. 1). Digital imaging has undergone an explosive development in the last decade, allowing access to high resolution and low-cost devices, especially in the case of 2D pictures.

Digital images used to collect data for GM analyses have to comply with some basic requirements, as imaging systems include artifacts related to acquisition, storage, and display processes [7], [21], [22], [23], [24]. The use of high-quality optical equipment and caution in specimen position and distance from the camera will help to reduce the effect of these artifacts.

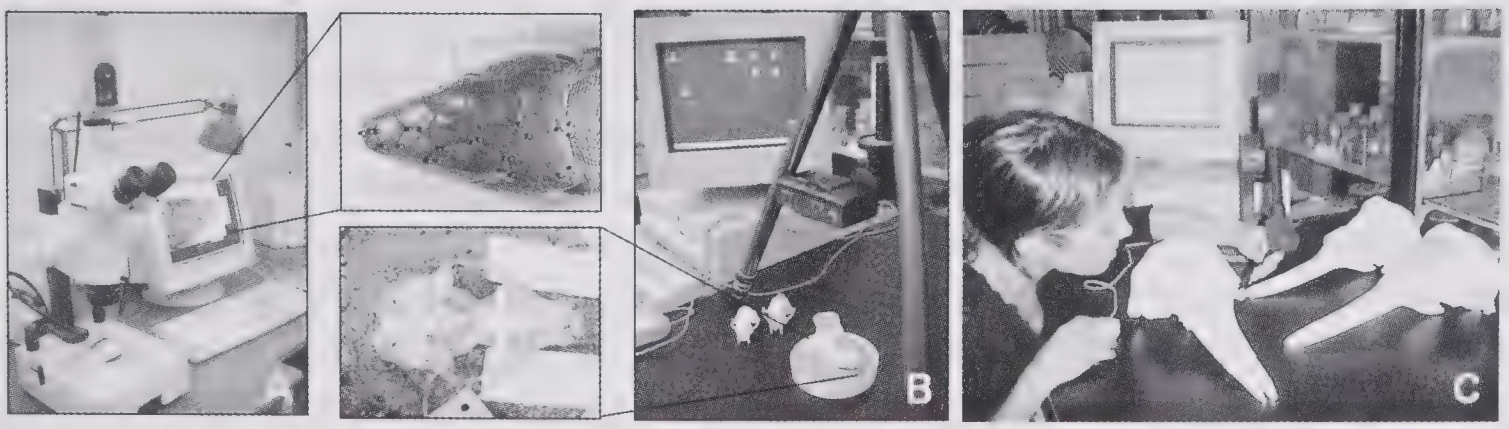


Fig. 1 – Devices used for 2D and 3D data recording for GM. **A**. Binocular microscope connected to a digital camera and a pc. **B**. Digital camera mounted on a tripod at a fixed distance from the skull. **C**. Microscribe used to get 3D coordinates directly from the skull.

The choice of digital camera should take into account resolution, accuracy, tonal range, color purity and accuracy, white balance, and image noise (see [23] and http://www.imaging-resource.com). Shadows and the reflecting surfaces of the object (bone surfaces often reflect a lot) may impede the view of important features like skull sutures. Lighting that maximizes the visibility of the whole object is recommended.

Once the equipment is calibrated, a standard protocol should be adopted for image recording that will retain all information needed for GM analyses (Fig. 2).

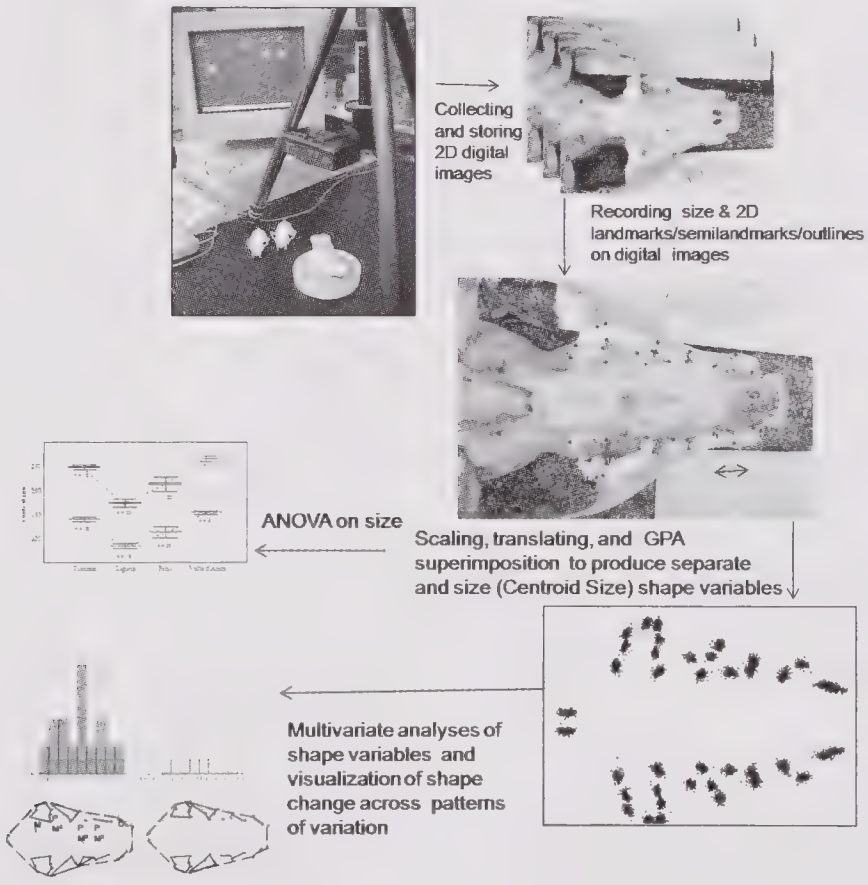

Collecting and storing 2D digital images
Recording size & 2D landmarks/semilandmarks/outlines on digital images
ANOVA on size
Scaling, translating, and GPA superimposition to produce separate and size (Centroid Size) shape variables
Multivariate analyses of shape variables and visualization of shape change across patterns of variation

Fig. 2 – Flow chart of 2D data recording and processing for GM.

The protocol should closely adhere to the following recommendations:

1. Place the object on a soft and possibly dark substrate to allow the object to be placed exactly horizontal to the plane.
2. Images should always include a scale factor and the specimen label.
3. Keep the object as much as possible in the centre of the image, and place the camera at such a fixed distance that distortion effects do not occur at image margins [24].

Three-dimensional images are more complicated, both in their nature and acquisition. They can be volumetric data encoding some spatial property, e.g., x-ray attenuation (CT scans) or water content (MRI-Magnetic field and Radiowave pulses), or they may consist of coordinates of point clouds representing the surface of a specimen (preferably with texture information). The production of such images has been prohibitively expensive, but is becoming increasingly affordable and accessible. Micro-CT scanners with micron-scale resolution are appearing within universities (e.g.http://www.ucalgary.ca/mousegenomics/3DMorphometrics, http://www.ctlab.geo.utexas.edu/index.php, http://micro-ct.at/). Usable surface scanning devices are ever more affordable (e.g., http://www.nextengine.com/). Students at FSU last year built a working scanner for only 15.00 USD (http://www.david-laserscanner.com/). Just as with 2D images, though, care must be taken to calibrate and test the product of any 3D scanning modality, and this information should be made available along with the image.

4 Images databanks, museums and GM analyses

Museum's collections play a fundamental role in geometric morphometric studies. By creating specific digital image databanks, museums would greatly speed up the data collection of large samples, reduce the costs of the research, and minimize damage to collections due to specimen handling. They could also benefit from extra income, as access to images could be regulated and charged, and advice from specialists could rapidly solve diagnostic and systematic problems often posed by specimen labels. To be suitable to GM analyses, image data banks should meet some basic requirements, including standards for image recording, acquisition, storage, and analysis of morphometric information. A number of excellent image repositories exist today. These include:

- Morphbank (http://www.morphbank.net/),
- MorphoBank (http://www.morphobank.org),
- and eSkeletons (http://www.eskeletons.org/),

designed primarily for two dimensional images, and:

- Aves3d (http://aves3d.org/),
- DigiMorph (http://digimorph.org/), and:
- the Open Research Scan Archive (http://plum.museum.upenn.edu/~orsa/ORSA/),

for (mainly) CT-based 3D images.

Such web-accessible repositories are valuable resources for any study concerned with the anatomical variation of the archived material, but there is much that could be improved. None of these (nor any other) repositories have incorporated into their structure the capacity for the direct acquisition, storage, and analysis of morphometric information. Neither do any of these provide the ability to easily interface with other archives. What is very much needed is an interface standard that would allow the query and retrieval of material from multiple online archives. Perhaps more practical, and maybe even better, would be the development of a meta-interface that could map the idiosyncrasies of the individual archives to a common access tool. Scientific research would be greatly enhanced, then, with even the basic capacity to search for and download images

(2- or 3D) and associated data for local analysis with standalone morphometric tools. Better still would be the direct support of such morphometric tools within the common interface and a secure, quality-controlled extensibility to the root archives to support morphometric annotation, e.g., labeled point (=landmark) coordinates, curve descriptors, etc.

As the growing use of medical imaging has allowed the production of 3D images of extant and fossil specimens, the analysis and visualization of 3D data and the combination of (semi-)landmarks, outlines, and surfaces are expected to yield a better description of changes in biological complexes. 3D analyses are affording several new perspectives in the field of human paleontology and physical anthropology (see for example [19], [20]). Progress is also expected in the study of covariation between subsets of landmarks [25] and the extension of landmark-based morphometrics to the analysis of articulated structures [26]. This progress would be greatly facilitated by standards-based, online archives that either directly incorporate or otherwise support the acquisition, storage, and analysis of morphometric data.

References

[1] J. J. Hanken and K. Hall (eds.), *The skull. Functional and evolutionary mechanisms*. vol. 3. Univ of Chicago Press, 1993.

[2] O. Thomas, "Suggestions for the nomenclature of cranial length measurements and the cheek teeth of mammals" *Proc.Biol.Soc.Wash.,* vol. 18, pp. 191-196, 1905.

[3] F. J. Rohlf and L. F. Marcus, "A revolution in morphometrics". *T.R.E.E,* vol. 8, pp. 129-132, 1993.

[4] F. L. A Bookstein, "Hundred years of morphometrics". *Acta Zool. Acad. Scient. Hungar.*, vol., 44, pp. 7-59, 1998.

[5] D. C. Adams, F. J. Rohlf and D. E. Slice, "Geometric Morphometrics: Ten Years of Progress Following the 'Revolution'" *Ital. J.Zool.*, vol. 71, pp. 5-16, 2004.

[6] D. E. Slice (ed.), *Modern Morphometrics in Physical Anthropology*. Kluwer Academic/Plenum, 2005.

[7] P. L. Forey and N. MacLeod (eds.), *Morphology, shape and phylogeny*. Syst. Ass. Sp. Vol. Ser. 64. London: Taylor and Francis, 2002.

[8] F. J. Rohlf "Geometric morphometrics and phylogeny". In: P. L. Forey and N. MacLeod (eds.), M*orphology, shape and phylogeny*. Syst. Ass. Sp., Vol. Ser. 64. London: Taylor and Francis pp. 175-193, 2002.

[9] A. Loy, "Morphometrics and theriological collections. Homage to Marco Corti". *Hystrix- It. J. Mammal.*, vol. 18, pp. 115-136, 2007.

[10] S. Elton and A. Cardini, "Anthropology from the desk? The challenges of the emerging era of data sharing*". J. Anthr.Sci.*, vol. 86, pp. 209-212, 2008.

[11] F. L. Bookstein "Size and shape spaces for landmark data in two dimensions". *Stat. Sci.,* vol. 1, pp. 181-222, 1986.

[12] D. E. Slice, "Landmark coordinates aligned by procrustes analysis do not lie in Kendall's shape space". *Syst. Biol.,* vol. 50, pp. 141-149, 2001.

[13] D. G. Kendall, "Shape-manifolds, Procrustean metrics and complex projective spaces". *Bull. Lond. Math. Soc.*, vol 16 pp. 81-121, 1984.

[14] D. G. Kendall, "Exact distributions for shapes of random triangles in convex sets". *Adv. Appl. Prob.*, vol. 17, pp. 308-329, 1985.

[15] F. L. Bookstein, *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, Cambridge, 1991.

[16] F. J. Rohlf, *tpsDig ver. 2.10*. Ecology and Evolution at SUNY Stony Brook, 2006.

[17] D. E. Slice, "Geometric morphometrics". *Ann.Rev. Anthr.,* vol. 36, pp. 261-281, 2007.

[18] P. Mitteroecker and P Gunz, “Advances in Geometric Morphometrics”. *Evol.Biol.* vol. 36, pp. 235-247, 2009.
[19] P. Gunz, P. Mitteroecker and F. L. Bookstein, “Semilandmarks in three dimensions”. In: D. E. Slice (*ed.). Modern Morphometrics in Physical Anthropology,* Kluwer, 2005.
[20] P. Gunz and K. Harvati, “The Neanderthal “chignon”: Variation, integration and homology. *J.f Hum. Evol.* vol. 52, pp. 262-274, 2007.
[21] F. J. Rohlf and F. L. Bookstein (eds.), *Proceedings of the Michigan morphometrics workshop*. Sp. Publ. N. 2. Univ. of Michigan Museum of Zoology, Ann Arbor, 1990.
[22] J. M. Becerra, E. Bello and A. Valdecasas “Building your own machine image system for morphometric analysis: a user point of view”. In: L. F. Marcus, E. Bello, A. Valdecasas (eds.), *Contribution to Morphometrics*, Monographias 8, Museo Nacional de Ciencias Naturales, pp. 66-92, 1993.
[23] A. Garcìa-Valdecasas, “Two-dimensional imaging: an update”. In: L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor and D. E. Slice (eds.), A*dvances in Morphometrics*, NATO ASI Series A: Life Sciences vol. 284. Plenum: New York: 71-81, 1996.
[24] M. L. Zelditch, D. L. Swiderski, H. D. Sheets and W. L. Fink, *Geometric Morphometrics for biologists: a primer*. London: Elsevier Academic Press, 2004.
[25] F. L. Bookstein, P. Gunz, P. Mitterocker, H. Prossinger, K. S Chafer and H. Seidler, “Cranial integration”. In *Homo:* singular warps analysis of the midsagittal plane in ontogeny and evolution. *J.Hum. Evol.*, vol. 44, pp. 167-187, 2003.
[26] D. C. Adams, “Methods for shape analysis of landmark data from articulated structures”. *Evol. Ecol. Res.*, vol. 1, pp. 959-970, 1999.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 249-250.
ISBN 978-88-8303-295-0. EUT, 2010.

Outline analysis for identifying *Limodorum* species from seeds

Sara Magrini, Sergio Buono, Emanuele Gransinigh, Massimiliano Rempicci, Silvano Onofri, Anna Scoppola

Abstract — *Limodorum trabutianum* Batt. is an orchid species of the Italian flora, with a central-western stenomediterranean distribution, that is sporadic in the western part of the distribution area of the more common *L. abortivum* (L.) Sw., an eurimediterranean species. It occurs in Italy only with a few populations in Tuscany, Latium, Umbria, Sicily and Sardinia [1], often with *L. abortivum* [2], [3], [4] from which it is easily recognizable only during anthesis for the denser inflorescence spike, the ribbon-like lip without differentiation in epychile and hypochile, and for the spur that is very short or absent [5]. On the contrary, the identification of these two *taxa* during the fruiting phase is rather difficult or even impossible. The aim of this study is to verify the taxonomic value of *Limodorum* seeds, particularly of their shape, as highlighted from recent studies for other orchids [6], [7], in order to establish its usefulness for recognizing the two species.
We have identified 5 Italian populations of the two *taxa*: 2 populations of *L. trabutianum*, one within the Marturanum Regional Park (Barbarano Romano, Viterbo), the other near Cortona (Arezzo), and 3 populations of *L. abortivum*, near S. Martino al Cimino (Viterbo), in the M. Casoli Reserve (Bomarzo, Viterbo), and in the same site of *L. trabutianum* within the Marturanum Park. The phenology of these populations was monitored to collect mature seeds from naturally dehiscing capsules. The intra- and interspecific variability of seed shapes was analyzed with the methodology of Elliptic Fourier descriptors [8], which allows to describe in terms of harmonics each two-dimensional shape with a closed outline. For this outline analysis we used the software package SHAPE 1.3 [9]. An average of 100 seeds from each species and from each site was photographed with a NIKON Coolpix 5000 camera mounted on a LEITZ-ARISTOPLAN microscope, obtaining 500 digital images with a resolution of 300 dpi and a size of 800 x 1000 pixels. All images were prepared using Adobe Photoshop 7.0: as a first step, every foreign element was eliminated from the picture, thereby isolating the single seed, then its contrast with the background was maximized, and finally all images were saved in .bps format (24bit). The color images were converted

S. Magrini, S. Onofri and A. Scoppola are with the Tuscia Germplasm Bank, Botanical Gardens of Viterbo - University of Tuscia, largo dell'Università s.n.c. - blocco C, Viterbo I 01100, Italy. E-mail of S. Magrini: magrini@unitus.it.
S. Buono, E. Gransinigh and M, Rempicci are with GIROS, Section "Etruria meridionale".

to binary with *Chain Coder* before tracing the outlines in Chain-code, a coding system that describes the geometrical information of the shapes. Then the Chain-code file was transformed into a Normalized Elliptic Fourier file using *Chc2Nef* using 20 harmonics. The matrix of the harmonic coefficients underwent a process of data normalization based on the first harmonic, to transform the data into shape variables. Subsequently, a PCA was performed on the variance-covariance matrix of normalized coefficients using *PrinComp*, which gives a graphical output of the principal components (average shape ± standard deviations).

The first results of the outline analysis confirm a low intraspecific variability of seed shape, but show a very high interspecific variability: *L. abortivum* seeds are very elongated, from fusiform to filiform, while *L. trabutianum* seeds are much wider and have a very lower length/width ratio. These results allow to distinguish between these two species even during the fruiting phase, simply using seed shape as a diagnostic character, avoiding the use of traditional morphometric analysis which need microscopic measurements.

Index Terms — image analysis, orchids, identification, plants.

References

[1] A. Scoppola and G. Spampinato (eds.), *Atlante delle specie a rischio di estinzione*, CD-ROM, 2005.
[2] R. Romolini and C. Merlini, GIROS Notizie, vol. 16, pp. 26-27, 2001.
[3] S. Buono and E. Gransinigh, GIROS Notizie, vol. 23, pp. 3-5, 2003.
[4] G. Ratini, GIROS Notizie, vol. 31, pp. 26-27, 2006.
[5] GIROS, *Orchidee d'Italia*, Il Castello, Cornaredo (MI), 303 pp., 2009.
[6] M. Aybeke, J. Pl. Biol., vol. 50, pp. 387-395, 2007.
[7] T. A. Akçin, Y. Ozdener and A. Akçin, J. Belgian. Bot., vol. 142, pp. 124-139, 2010.
[8] F. P. Kuhn and C. R. Giardina, Comp. Graph. Ima. Proc., vol. 18, pp. 236-258, 1982.
[9] H. Iwata, Y. Ukai and J. Heredity, vol. 93, pp. 384-385, 2002.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 251-256.
ISBN 978-88-8303-295-0. EUT, 2010.

Geometric morphometrics as a tool to resolve taxonomic problems: the case of *Ophioglossum* species (ferns)

Sara Magrini, Anna Scoppola

Abstract — A modern method, geometric morphometrics, was used to clarify the taxonomic position of the European *Ophioglossum* species: *O. azoricum*, *O. lusitanicum*, and *O. vulgatum*. The identification of these *taxa* by traditional methods is rather difficult, due to different taxonomic interpretations. Sterile leaf shapes were investigated using a landmark-based method and the Fourier analysis of outlines. Both methods highlight the shape and the base of the leaf as an important diagnostic character.

Index Terms — geometric morphometrics, landmark, *Ophioglossum*, outline analysis.

1 Introduction

Geometric morphometrics [1], [2] is a technique for multivariate shape analysis that preserves the integrity of biological shape, avoiding its reduction to linear and angular measures that do not include information concerning the geometric relationships of the entire subject. allowing to identify the anatomical areas of morphological remodelling. Two different kinds of geometric morphometrics are most widely used: landmark-based methods such a Thin Plate Splines analysis [3], and Fourier analysis of outlines [4]. The theoretical formulation has been developed in recent decades by the synthesis of multivariate analysis methods of covariance matrices and methods for direct visualization of changes in the biological shape [1], [2], [5]. Multiple applications exist in the biomedical field, in paleontology, anthropology and zoology. In contrast, geometric morphometrics was almost absent from botanical research until a few years ago [3], with a few exceptions for *Dactylorhiza* [6], *Quercus* [7], and other groups.

S. Magrini is with the Tuscia Germplasm Bank, University of Tuscia, Viterbo, I 01100. E-mail: magrini@unitus.it.
A. Scoppola is with the Viterbo Botanical Gardens, University of Tuscia, Viterbo, I 01100. E-mail: scoppola@unitus.it.

Many groups of plants still require more intensive taxonomic studies. One of them is the genus *Ophioglossum* C. Presl, an ancient fern genus of about 25-30 species of *Ophioglossaceae*, with a cosmopolitan but primarily tropical and subtropical distribution. The simple morphology of the sporophyte - a single leaf or a few - has limited the number of characters available for building classifications on morphological ground and clarifying relationships, forcing investigators to rely on details of frond size or sporangia number as diagnostic markers. Problems associated with the evaluation of these often subtle differences gave rise to different taxonomic interpretations of the genus [8]. Furthermore, the number of taxonomic studies of *Ophioglossum* species is relatively low in Europe [9], [10], so that many questions concerning *Ophioglossum azoricum* C. Presl and its relationships with the other species (*O. vulgatum* L. and *O. lusitanicum* L.), remain unresolved. Our study aims at testing a modern method - geometric morphometrics - for resolving some of these problems and clarifying the taxonomic position of these *Ophioglossum* species.

2 Materials and Methods

2.1 Collecting materials

We have analysed three species: *O. azoricum*, *O. lusitanicum*, and *O. vulgatum*, as well as some critical populations from Tuscany (Moriglion di Penna and Monte Pianello, Lucca) and Latium (Selva del Lamone, Viterbo). We used 100 *exsiccata* from 9 European herbaria. We photographed with a digital camera all the sterile leaves in the specimens of Cagliari (CAG), Florence (FI), Genoa (GE), Pisa (PI), Siena (SIENA), Viterbo (UTV), and Valencia (VAL), and we collected images from the website of Paris (P), and Vienna (W). We obtained over 203 images, standardized with a resolution of 300 dpi and a size of 800x1000 pixels, prepared using *Adobe Photoshop 7.0.* As a first step, every foreign element was eliminated from the picture, thereby isolating the sterile leaf, and then the contrast with background was maximized.

2.2 Landmark-based method

The *TPS - Thin-Plate Splines* software package was used for landmark-based morphometrics analysis: *TpsDig2.12* to digitize landmarks [11] and *TpsRelw 1.42* for statistical analysis [12]. In our case, we could find only 3 homologous points (landmarks), due to the shape of the leaf, with an entire margin, and the lack of evident veining. A grid of 11 equally spaced horizontal lines was superimposed on all images, taking care to position it perpendicular to the longitudinal axis of the frond and to align the apex and the base with the top and the bottom line of the grid, respectively, to digitize other 18 semi-landmarks (points defined relatively to other landmarks). The next step was the development of matrix of landmark coordinates through *TpsRelw 1.42*. This program operates through an algorithm used to describe the mechanical deformation of thin metal foils (thin plate spline or TPS), building linear combinations of the landmark coordinates.

The new variables - called "shape variables" - allow the creation of deformation grids that visually illustrate the change in shape compared to the average shape of the sample (*consensus*). We have analysed the entire dataset and the partial datasets corresponding to each species and population to defy inter- and intraspecific variability, respectively, and we have performed a Relative Warps Analysis on the covariance matrix.

2.3 Outline Analysis

For Outline Analysis we used the software package *SHAPE 1.3* [13]. based on the methodology of Elliptic Fourier descriptors, which allows to describe - in terms of harmonics - each type of two-dimensional shape with a closed outline. All images were saved in .bps format (24bit) and were converted to binary with *Chain Coder* before tracing the outlines in *Chain-code*, a coding system that describes the geometrical information on the shapes. Then, the *Chain-code* file was transformed into a Normalized Elliptic Fourier file with *Chc2Nef*, using 20 harmonics on 77 components. The matrix of the harmonic coefficients underwent normalization based on the first harmonic, to transform the data into shape variables. Subsequently, a PCA was performed on the variance-covariance matrix of normalized coefficients (Elliptic Fourier Descriptors) using *PrinComp*, that gives a graphical output of the average shape ± the standard deviation.

3 Results

We report the results of the Relative Warps Analysis (RWA) on the entire dataset. The analysis of the first two RWs (accounting for 85.10% and 6.07% of the total variance, respectively), shows a clear separation between *O. vulgatum* and *O. lusitanicum* along RW1, while in the centre of the plot several samples of *O. azoricum* are very close to *O. lusitanicum*. The samples from Lamone (LAM) and M. Pianello (PIA) overlap with *O. azoricum*, while those from Moriglion di Penna (MOR) are distributed along RW1 in the areas of *O. vulgatum* and of *O. azoricum*. Moving along RW1 from negative values, corresponding to *O. lusitanicum*, to the positive values of *O. vulgatum*, the associated deformation grids show a vertical contraction and an expansion in width, particularly in the proximal part of the leaf (Fig. 1).

The results of the Outline Analysis of the entire dataset (Fig. 2a) are consistent with those from RWA: PC1 (accounting for 88.33%) shows the 2 extreme shapes diversified in the proximal part of the leaf, corresponding to *O. lusitanicum* for the negative end, to *O. vulgatum* for the positive end. The results for the single species show a lower variability in *exsiccata* of *O. azoricum "typus"* (PC1=56.15%) (Fig. 2b), and of *O. lusitanicum* (PC1=61.34%) (Fig. 2e), while the greater variability is in *O. vulgatum* (PC1=70.67%) (Fig. 2d), confirming the difficulty of a correct identification by traditional morphological methods due to the different taxonomic interpretations.

The results were used to identify the critical samples from LAM, PIA and MOR. The LAM outlines (Fig. 2f) have a low variability (PC1=64.15%), they are akin to *O. azoricum* outlines (Fig. 2b-c), as the PIA samples (Fig. 2g), which have a

greater variability (PC1=80.67%). On the contrary, the MOR samples (Fig. 2h) are diversified (PC1=91.54%), showing 2 extreme shapes, corresponding to *O. vulgatum* (Fig. 2d) and *O. azoricum* outlines (Fig. 2 b-c).

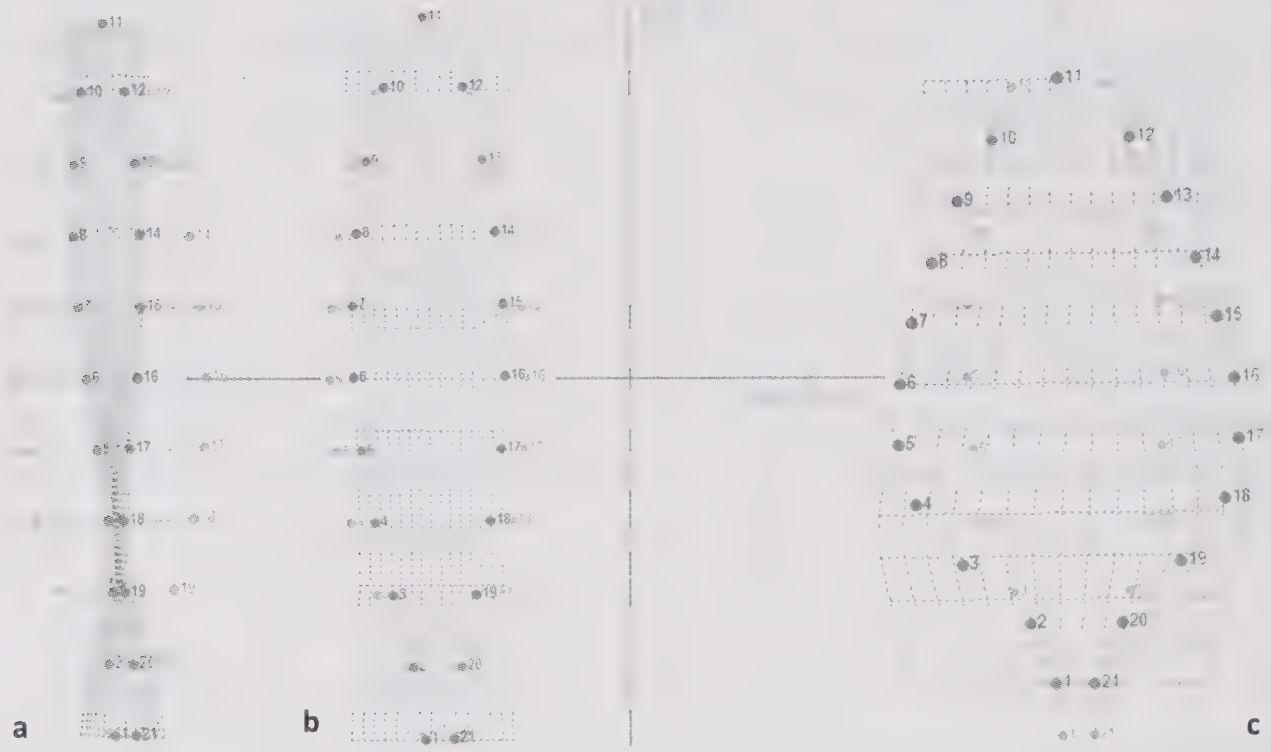


Fig. 1 – Deformation grids along the first Relative Warp with vectors of deformation relative to average shape of fronds, showing the anatomical areas of morphological remodelling.

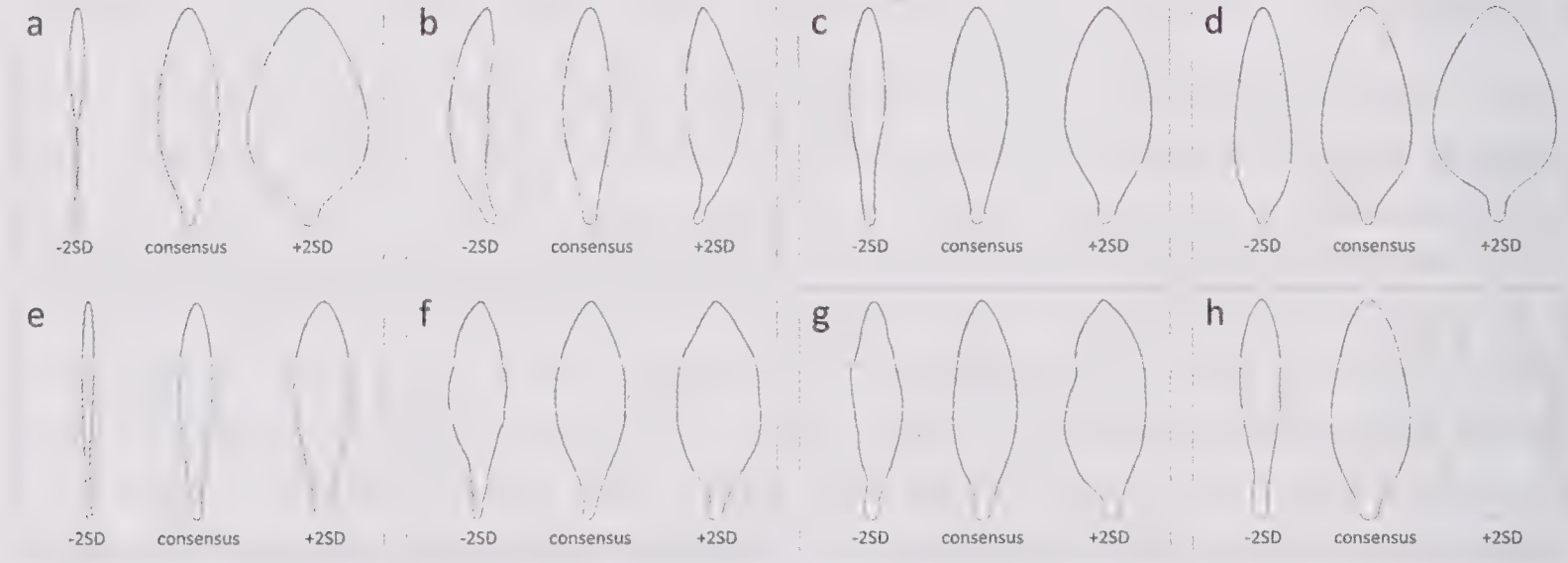


Fig. 2 – Results from Outline Analysis. First principal component of: (a) the entire dataset, PC1=88.33%; (b) *O. azoricum typus*, PC1=56.15%; (c) *O.azoricum*, PC1=64.98%; (d) *O. vulgatum*, PC1=70.67%; (e) *O. lusitanicum*, PC1=61.34%; and *Ophioglossum sp.* from: (f) LAM, PC1=64.15%; (g) PIA, PC1=80.67%; (h) MOR, PC1=91.54%.

4 Discussion

The geometric morphometrics analysis gave results that lead to a better characterization of the 3 *Ophioglossum* species.

The landmark-based analysis does not completely discriminate among the three species. It clearly separates *O. lusitanicum* from *O. vulgatum*, but *O. azoricum* and *O. lusitanicum* are widely overlapping.

More information derives from the deformation grids that highlight a diagnostic character like the relative width of the leaf, especially in its proximal part, as

confirmed by the Outline Analysis. The graphical outputs of the PCA of outlines show the differences between the three species, confirming shape and base of the leaf as the main diagnostic characters. *O. vulgatum* is well distinct from the other two species by the shape of the lamina, which is from lanceolate to broadly ovate with a large round and attenuated base. *O. azoricum* and *O. lusitanicum* have a more or less cuneate base, but they are differentiated in the leaf shape: from lanceolate to narrow ovate in *O. azoricum*, lanceolate-linear in *O. lusitanicum*. Geometric morphometrics also allows to identify the critical samples: LAM and PIA samples can be referred to *O. azoricum*, as confirmed by both Relative Warp and Outline Analysis. Instead, the Outline Analysis of MOR samples shows a higher variance, although this could be attributed to the variability of *O. vulgatum;* probably in this site *O. azoricum* and *O. vulgatum* coexist.

5 Conclusion

The study led to a more accurate characterization of the three species, with the identification of valid diagnostic characters, and the presumable discovery of 2 new sites of *O. azoricum* in Italy, M. Pianello (Lucca) and Selva del Lamone (Viterbo), even if other analyses (e.g. molecular and caryological) are suggested. The research will continue with the revision of all samples that showed abnormal or out-of-range values, and by increasing the dataset with *exsiccata* from other herbaria. These results also will allow understanding and defining the actual distribution of *Ophioglossum* species in Italy and Europe.

Acknowledgement

The authors wish to thank the curators of the cited herbaria and dr. L. Peruzzi (University of Pisa) for providing images or exsiccata for this study.

References

[1] F. Bookstein, *Morphometric Tools for Landmark Data*. Cambridge Univ. Press, 1991.

[2] F. J. Rohlf and L. F. Marcus, "A Revolution in Morphometrics". *Trends Ecol. Evol.*, vol. 8, pp. 129-132, 1993.

[3] D. C. Adams, F. J. Rohlf and D. E. Slice, "Geometric Morphometrics: Ten Years of Progress Following the 'Revolution'". *Ital. J. Zool.*, vol. 71, pp. 5-16, 2004.

[4] R. J. Jensen, K. M. Ciofani and L. C. Miramontes, "Lines, Outlines and Landmarks: Morphometric Analyses of Leaves of *Acer rubrum*, *Acer saccharinum* (*Aceraceae*) and Their Hybrid". *Taxon*, vol. 51, pp. 475-492, 2002.

[5] L. R. Rabello, B. Bordin and S. Furtado, "Shape Distances, Shape Spaces and the Comparison of Morphometric Methods". *Trends Ecol. Evol.*, vol. 15, pp. 217-220, 2000.

[6] A. B. Shipunov and R. M. Bateman, "Geometric Morphometrics as a Tool for Understanding *Dactylorhiza* (Orchidaceae) Diversity in European Russia". *Biol. J. Linnean Soc.,* vol. 85, no. 1, pp. 1-12, 2005.

[7] V. Viscosi, O. Lepais, S. Gerber and P. Fortini, "Leaf Morphological Analyses in Four European Oak Species (*Quercus*) and Their Hybrids: a Comparison of Traditional and Geometric Morphometric Methods", *Plant Biosyst.*, vol. 143, no. 3, pp. 564-574, 2009.

[8] W. D. Hauk, C. R. Parks and M. W. Chase, "Phylogenetic Studies of Ophioglossaceae: Evidence from rbcL and trnL-F Plastid DNA Sequences and Morphology". *Mol. Phyl. Evol.*,

vol. 28, pp. 131-151, 2003.
[9] A. M. Paul, "The Status of *Ophioglossum azoricum* (Ophioglossaceae: Pteridophyta) in the British Isles". *Fern Gaz.*, vol. 13, pp. 173-187, 1987.
[10] L. Peruzzi, G. Cesca and D. Puntillo, "*Isoëtes* (Isoetaceae), *Ophioglossum* and *Botrychium* (Ophioglossaceae) in Calabria (S Italy): More Karyological and Taxonomical Data". *Caryologia*, vol. 56, no. 3, pp. 355-359, 2003.
[11] F. J. Rohlf, *TpsDig2, Digitize Landmarks and Outlines, Version 2.12*. Dep. of Ecology and Evolution, State University of New York at Stony Brook, 2007.
[12] F. J. Rohlf, *TpsRelw, Version 1.42*. Dep. of Ecology and Evolution. State University of New York, Stony Brook, 2005.
[13] H. Iwata and Y. Ukai, "SHAPE: a Computer Program Package for Quantitative Evaluation of Biological Shapes Based on Elliptic Fourier Descriptors.", *J. Heredity*, vol. 93, pp. 384-385, 2002.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 257-261.
ISBN 978-88-8303-295-0. EUT, 2010.

Geometric morphometric analysis as a tool to explore covariation between shape and other quantitative leaf traits in European white oaks

Vincenzo Viscosi, Anna Loy, Paola Fortini

Abstract — In this study, 2-Block Partial Least-Squares analysis was used to explore the pattern of covariation between two sets of leaf morphological data (landmarks and traditional measures), sampled on 273 trees in a mixed forest of *Quercus frainetto, Q. petraea* and *Q. pubescens,* in southern-central Italy. Two groups of related variables were detected and the three oak species were highly significant discriminated by CVA computed on dimensions extracted by 2B-PLS analysis. *Q. frainetto* was characterized by obovate leaf blade with a short petiole and leaves were greater than in the other species. *Q. petraea* was differentiated by acute basal and apical regions, while *Q. pubescens* has higher values of leaf compactness, pubescence and length of trichomes. The high degree of classification accuracy of this combined approach advocates its extension to other problematic species and highlights its importance as an exploratory tool in plant ecology, physiology and taxonomy.

Index Terms — geometric morphometrics, leaf shape, Quercus, Two-Block Partial Least-Squares.

1 Introduction

The morphological variability in subgenus *Quercus* has been studied for a long time. Several analyses, usually based on leaf morphological data, supported the existence of high variability within and among these oak species ([1], [2], [3], [4], [5]). Here a landmark based geometric morphometric approach is used to analyse the leaf morphological pattern of covariation

V. Viscosi is with the Museo Erbario del Molise, Department STAT, University of Molise, Pesche (IS) IT-86090. E-mail: vincenzo.viscosi@unimol.it.
P. Fortini is with the Museo Erbario del Molise, Department STAT, University of Molise, Pesche (IS) IT-86090. E-mail: fortini@unimol.it.
A. Loy is with Department STAT, University of Molise, Pesche (IS) IT-86090. E-mail: vincenzo.viscosi@unimol.it. E-mail: loy@unimol.it.

between shape and other traditional traits in three white oak species. We used Two-Block Partial Least-Squares analysis (2B-PLS) [6] (i) to summarize the greater part of leaf phenotypic variability in few linear dimensions, (ii) to detect the degree of differentiation among species, (iii) to obtain a useful classification method for unidentified specimens.

2 Materials and Methods

2.1 Plant material and data sampling

A total 273 oaks were sampled in a woody community (Italy, Molise) hosting three hybridizing white oak species. Species assignment followed Schwarz (1996): *Q. petraea* (n=122), *Q. pubescens* (n=57), *Q. frainetto* (n=56), hybrids (38). Ten leaves for each tree were photographed for morphological analyses. Eleven landmarks and 25 traditional morphological characters were collected on each leaf [5]. Eight morphological characters (area, perimeter, roundness, elongation, feret diameter, compactness, major axis length, minor axis length) were automatically recorded on digital images by means of ImageTool ver. 5.1 [7]; five variables were recorded manually by a digital calliper: lamina length (LL), petiole length (PL), lobe width (LW), sinus width (SW), length of lamina at largest width (WP), number of lobes (NL) and number of intercalary veins (NV). Five variables were obtained by ratios between previously measured variables [1]: lamina shape (OB), petiole ratio (PR), lobe depth ratio (LDR), percentage venation (PV), lobe width ratio (LWR); four variables were observed and recorded as ordinal variables ([1], [8],): basal shape (BS), pubescence of abaxial surface, adaxial surface and petiole. Finally, the length of hairs was measured [9].

2.2 Statistical analysis

For each tree, the mean of 25 leaf variables and the landmark consensus configuration was computed. 2B-PLS analysis was used to investigate the covariation pattern among the 11 shape descriptors and the 25 traditional variables. Linear combinations of partial warps and other quantitative traits were visualized as deformation grids through tpsPLS 1.18 [10].

Finally, the linear dimensions, derived from 2B-PLS analysis, were analysed by Canonical Variate Analysis (CVA) to detect the degree of species differentiation among species and to obtain a classification rule for unidentified specimens.

3 Results

The first two significant dimensions (2B-PLS) accounted for 95.34% and 2.96% of total covariance, respectively. The correlations between shape descriptors and morphological variables for the two first dimensions were 0.884 ($p < 0.001$) and 0.780 ($p < 0.001$), respectively.

The ordination plot showed a clear separation of *Q. frainetto, Q. pubescens*

and *Q. petraea* (Fig. 1): *Q. frainetto* is clearly distinguished from both *Q. petraea* and *Q. pubescens*, while the two latter partially overlapped. Moreover, two main groups of variables that covariated in relation to leaf shape were detected (Fig. 2). Weight along these dimensions in relation to each deformation grid are shown in Fig. 3.

Linear dimensions derived from 2B-PLS analysis were subjected to CVA (Fig. 4): the two CVs explained 76.8% (Wilks'λ = 0,006; df = 22; p<0.0001) and 23,2% (Wilks'λ = 0,136; df = 10; p<0.0001) of total variability, respectively. All three species were highly significant discriminated. The cross validation confirmed the result and 100% of specimens were correctly classified.

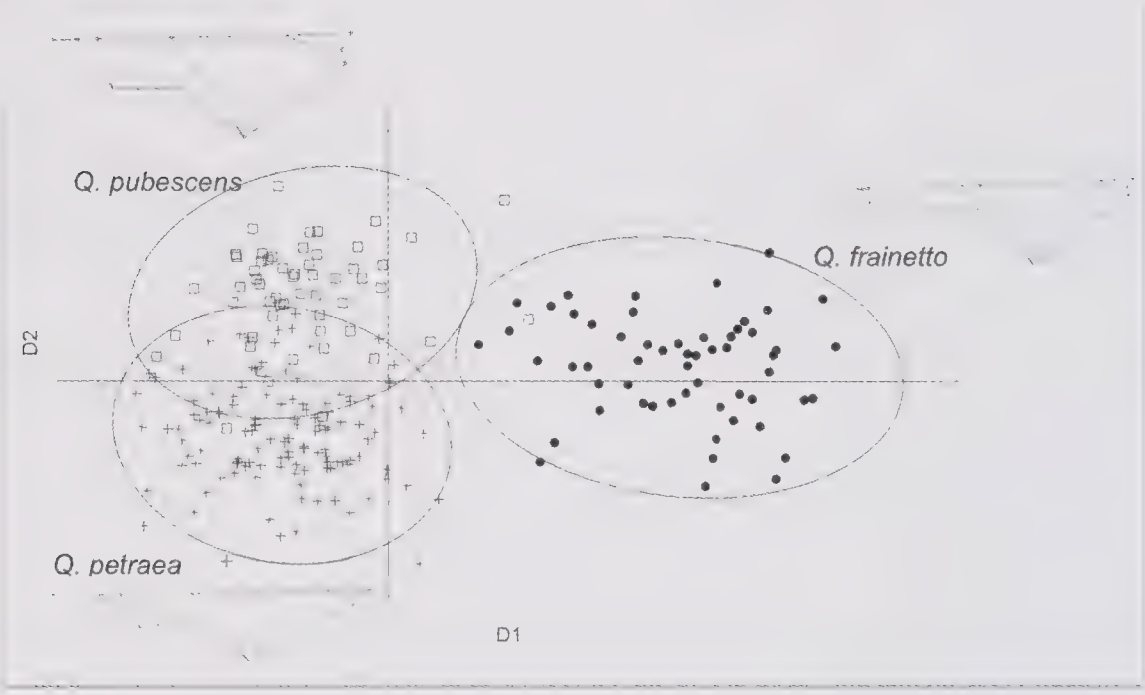


Fig. 1 – Ordination plot for the projections of the pure species onto first two significant dimensions. Filled circles = *Q. frainetto*; crosses = *Q. petraea*; empty squares = *Q. pubescens*. Hybrids are not shown.

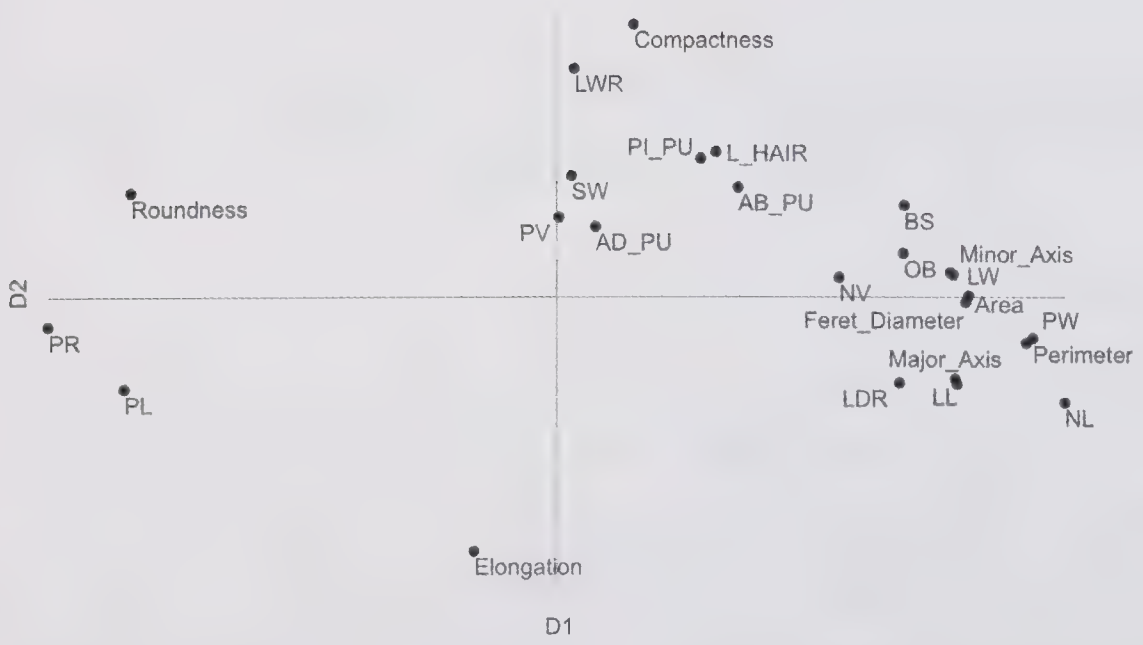


Fig. 2 – Loadings plot of 25 morphological variables onto the first two significant dimensions.

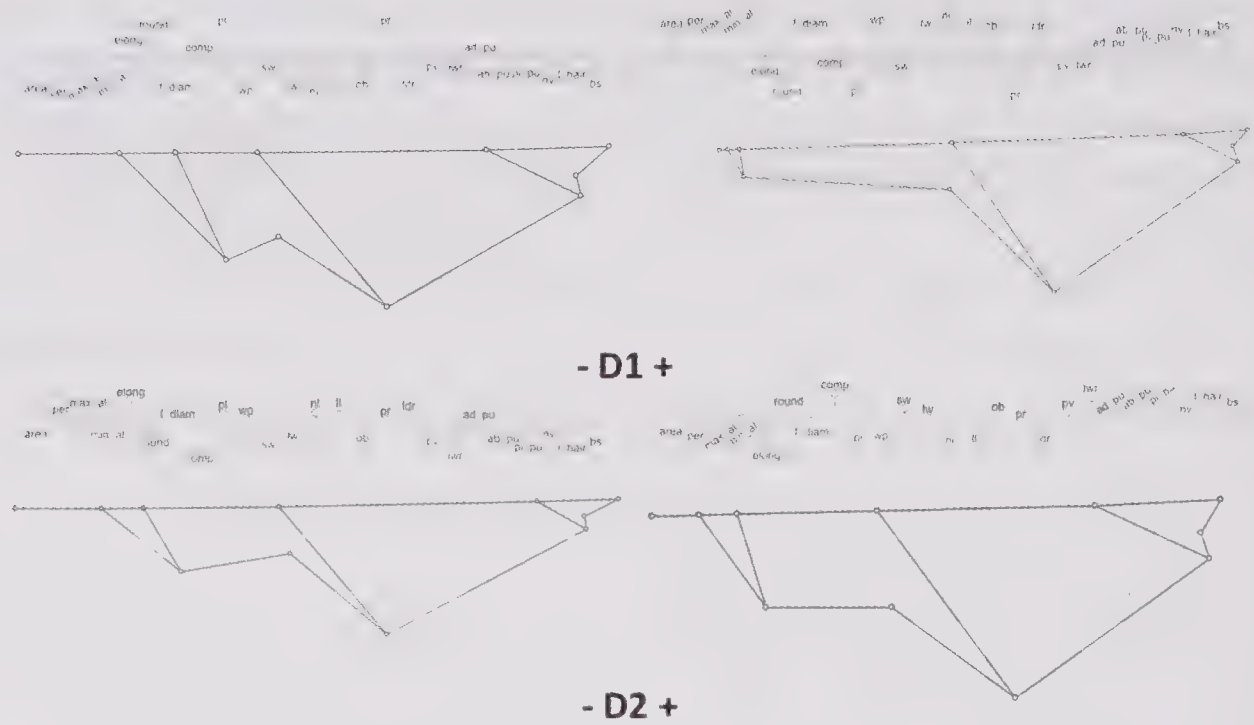


Fig. 3 – Shape variation of leaves along the first two significant dimensions of 2B-PLS. Deformation grids are shown for the negative (left) and positive (right) extremes of D1 (top) and D2 (bottom). The bar-plot show the degree of covariation between shape and the 25 traditional characters (see text for codes).

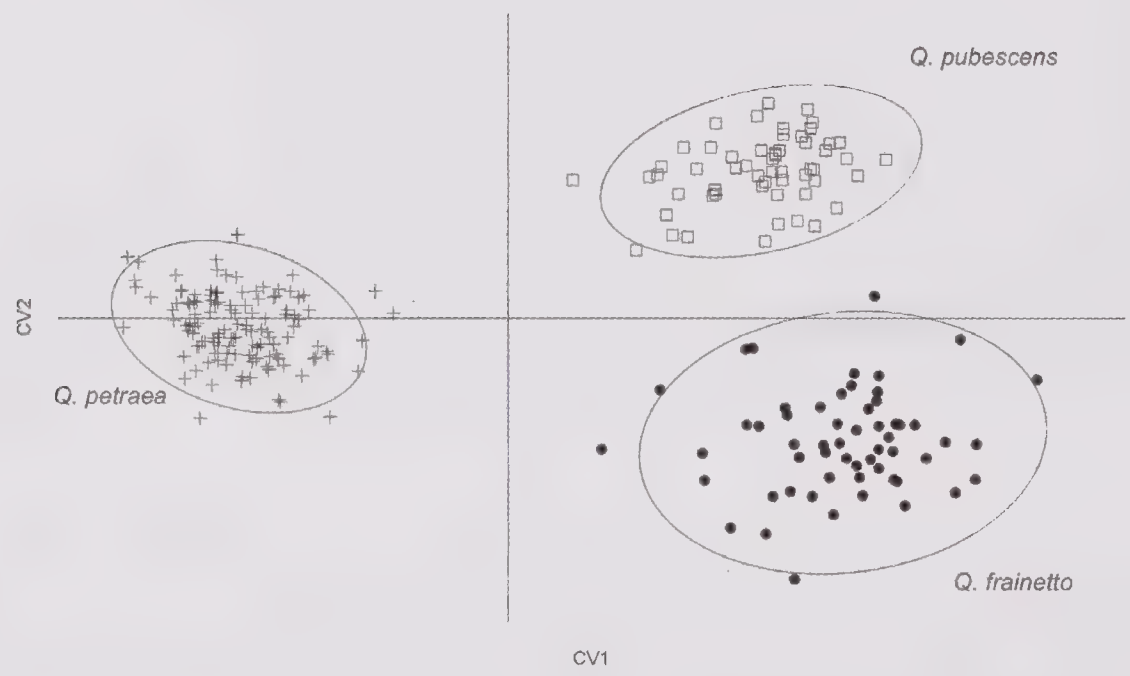


Fig. 4 – Ordination plot of the pure species onto the first two canonical variates. Filled circles = *Q. frainetto*; crosses = *Q. petraea*; empty squares = *Q. pubescens*. Hybrids are not shown.

4 Conclusion

This approach results as a useful tool to analyze traditional morphological variables in relation to leaf shape, one of the most important visual attributes to characterize white oak species, playing an important role in pattern of recognition, and in the systematics of this plant subgenus.

2B-PLS results allowed to define a typical leaf shape for each oak species, and to associate this shape to other morphological traditional features.

Results confirmed the high diagnostic power of the geometric morphometric approach. Moreover, the visualization of the leaf shape differences, associated

to other groups of correlated morphological traits, allowed to obtain a clear diagnosis of leaf morphology for each species.

Q. frainetto was characterized by grater leaves and short petiole, obovate leaf blade and high deepness of lobes. *Q. petraea* was characterized by a longer petiole, more acute basal and apical regions, and a more deeply lobed lamina than *Q. pubescens,* which has higher values of leaf compactness, pubescence and length of trichomes.

The high degree of classification accuracy of this combined approach advocates its extension to other problematic species and highlights its importance as an exploratory tool in plant ecology, physiology and taxonomy.

References

[1] A. Kremer, L. J. Dupouey, J. D. Deans, J. Cottrell, U. Csaikl, R. Finkeldey, S. Espinel, J. Jensen, J. Kleinschmit, B. Van Dam, A. Ducousso, I. Forrest, U. L. de Heredi, A. J. Lowe, M. Tutkova, R. C. Munro, S. Steinhoff and V. Badeau, "Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands". *Ann. Sci. For.,* vol. 59, pp. 777-787, 2002.

[2] P. Bruschi, G. Vendramin, F. Bussotti and P. Grossoni, "Morphological and Molecular Differentation between *Quercus petrea* (Matt.) Liebl. and *Quercus pubescens* Willd. (Fagaceae) in Northern and Central Italy". *Ann. Bot.,* vol. 85, pp. 325-333, 2000.

[3] S. Ponton, J.-L. Dupouey, E. Dreyer, "Leaf morphology as species indicator in seedlings of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl.: modulation by irradiance and growth flush". *Ann. For. Sci.*, vol. 61. pp. 73-80, 2004.

[4] D. Uribe-Salas, C. Sáenz-Romero, A. González-Rodríguez, O. Téllez-Valdéz and K. Oyama, "Foliar morphological variation in the white oak *Quercus rugosa* Née (Fagaceae) along a latitudinal gradient in Mexico: Potential implications for management and conservation". *For. Ecol. Manag.*, vol. 256, pp. 2121-2126, 2008.

[5] V. Viscosi, P. Fortini, D. E. Slice, A. Loy and C. Blasi, "Geometric morphometric analyses of leaf variation in four oak species of subgenus *Quercus* (Fagaceae)". *Plant Biosyst.,* vol. 143, pp. 575-587, 2009.

[6] F. J. Rohlf and M. Corti, "Use of two-block partial least-squares to study covariation in shape". *Syst Bio.,* vol. 49, pp. 740–753, 2000.

[7] S. B. Dove, "UTHSCSA ImageTool program 3.0", 2000.

[8] P. Kissling, "Les poils des quatre espèces de chênes du Jura (*Quercus pubescens, Q. petraea, Q. robur* et *Q. cerris*)". *Ber. Schweiz. Bot. Ges.*, vol. 87, pp. 1-18, 1977.

[9] P. Bruschi, G. G. Vendramin, F. Bussotti and P. Grossoni, "Morphological and molecular diversity among Italian populations of *Quercus petraea* (*Fagaceae*)". *Ann. Bot.,* vol. 91, pp. 707-716, 2003.

[10] F. J. Rohlf, "TpsPLS 1.18". Department of Ecology and Evolution, State University New York. Stony Brook, 2006.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 263-268.
ISBN 978-88-8303-295-0. EUT, 2010.

Landmark based morphometric variation in Common dolphin (*Delphinus delphis* L.,1758)

Paola Nicolosi, Anna Loy

Abstract — In this study we compare Mediterranean stocks of *Delphinus delphis* (L., 1758) with other populations of the same species coming from different seas using a geometric morphometrics method. The aim is to define the patterns of geographical variation of *Delphinus delphis* through a geometric morphometrics analysis of the skulls of 124 individuals from seven marine areas (West and East Pacific Ocean; North-east and South-East Atlantic Ocean, West and East Indian Ocean, Mediterranenan Sea).

Index Terms — common dolphin, geometric morphometrics, skulls.

1 Introduction

The Mediterranean sea has experienced significant changes in the last decades in terms of biodiversity, due to a combination of environmental and anthropogenic influences. In this project we focus the attention on the common dolphin, *Delphinus delphis*, whose Mediterranean population was drastically reduced starting from the Sixties and is considered as "Endangered" from 2003 "[3], [4]". Analyses where devoted to clarify the pattern of geographic variation of the species through a geometric morphometric approach, and to evaluate any specific differentiation/adaptation of the Mediterranean stock with respect to other populations across the range of the species. Due to the difficulties related to data collection and records in the field, the museum collections represented the primary source of information, as in many other Cetacea"[6]".

P. Nicolosi is with the Museo di Zoologia, Università degli Studi di Padova, 35121 Padova, Italy. E-mail: paola.nicolosi@unipd.it.
A. Loy is with the Dipartimento di Scienze e Tecnologie per l'Ambiente e il Territorio, Università degli Studi del Molise, Pesche (IS), Italy. E-mail: a.loy@unimol.it.

2 Materials and Methods

A total 124 skulls of adult specimens from seven marine areas across the distribution range of the species (Tab. 1) were photographed on dorsal, ventral and lateral projections with a digital camera using a standard procedure to avoid the effects of distortion. Previous analyses on the absence of sexual dimorphism in the shape of the skull "[9]" allowed to pool males and females.

The analysis of 24 two-dimensional cartesian coordinates (landmarks) have been recorded on the various projections using the software tpsDig "[10], [11]" (Fig. 1). Data have been translated, rotated and superimposed through a General Procrustes Analysis, GPA "[11]" using the tpsRelw software "[10]". Centroid sizes were stored for allometric and size variation evaluations. Multivariate ordination of specimens was performed through Relative Warp Analysis on the weight matrix of aligned specimens.

SEAS	*REGION*	*LOCATION*	*SAMPLES*	*SYMBOL*
Mediterranean	-	-	14	
Atlantic	*North East*	Portugal	26	
		Danmark, Netherlands	7	
		Ireland	5	
	South East	East Africa	31	
		South Africa	1	
Indian	*West*	Oman	3	
		Persian Gulf	1	
	East	Australia	3	
Pacific	*South West*	New Zealand	26	
	South East	Perù	7	
TOT			**124**	

Tab. 1 – Number of specimens and their geographical location: 14 from the Mediterranean sea (Italians Naturalistic and Zoological Museums), 70 from the Atlantic Ocean (Lisbon Natural History Museum, 26; Zoological Museum of Amsterdam, 25; Zoological Museum of Copenhagen, 19), 33 from the Pacific Ocean (Zoological Museum of Amsterdam, 31; Zoological Museum of Copenhagen, 2), 7 from the Indian Ocean (Zoological Museum of Amsterdam, 3; Zoological Museum of Copenhagen, 4.

3 Results

Fig.2 shows the results of ordination analysis of the residual from GPA for the dorsal view of the skulls, while Fig.3 shows the results of classification analysis run on the Mediterranean, Atlantic, Indian, and Pacific stocks. The first two PC (retaining 37,7% and 10,5% of cumulative variation respectively) do not allow a clear identification of different stocks except for the Indian ocean

sample. Nevertheless Mahalanobis distances among groups derived from CVA scores are highly significant (Tab. 2). Procrustes distances among populations confirm the Indian stock as the most divergent from all other samples, while the Mediterranean is the most different with respect the Atlantic and the Pacific dolphins.

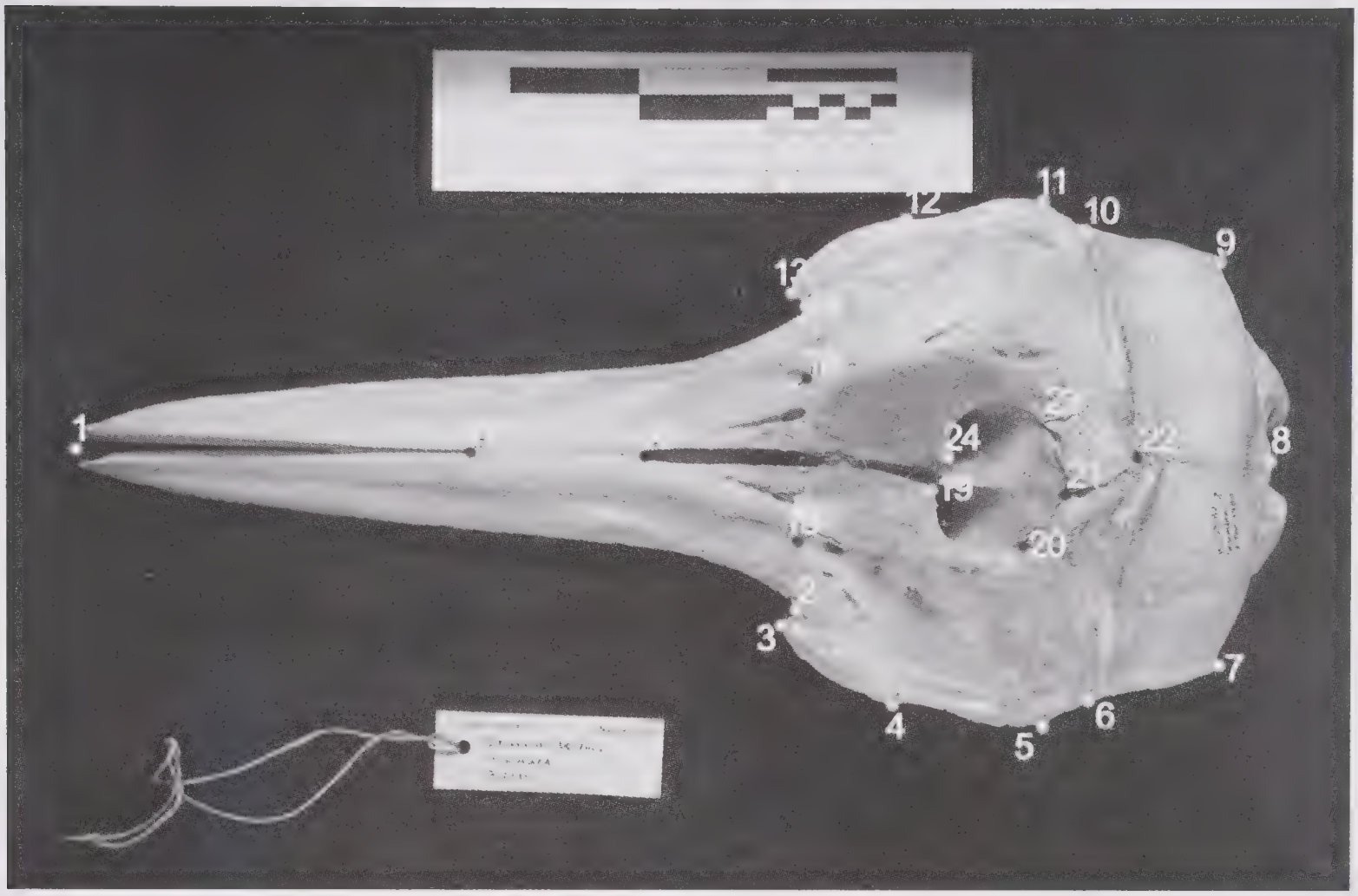


Fig.1 – Landmark recorded on the dorsal projection of the skull.

	Atlantic	Indian	Mediterranean	Pacific
Atlantic	-	0,0552	0,0211	0,0220
Indian	5,0605***	-	0,0521	0,0440
Mediterranean	2,7085***	5,2529***	-	0,0247
Pacific	3,0821***	5,6897***	3,2302***	-

Tab. 2 – Above diagonal: Procrustes distances among populations; below diagonal: Mahalanobis distances among groups derived from CVA scores. *** $P < 0.0001$.

The deformation grid on the left in the graph (Fig.2) is referred to the shape changes characterizing the Indian Ocean dolphins. The skulls of these specimens shown an elongation of intermaxilla bones and infraorbita foramina aligned to the antorbital notch respect the mean.

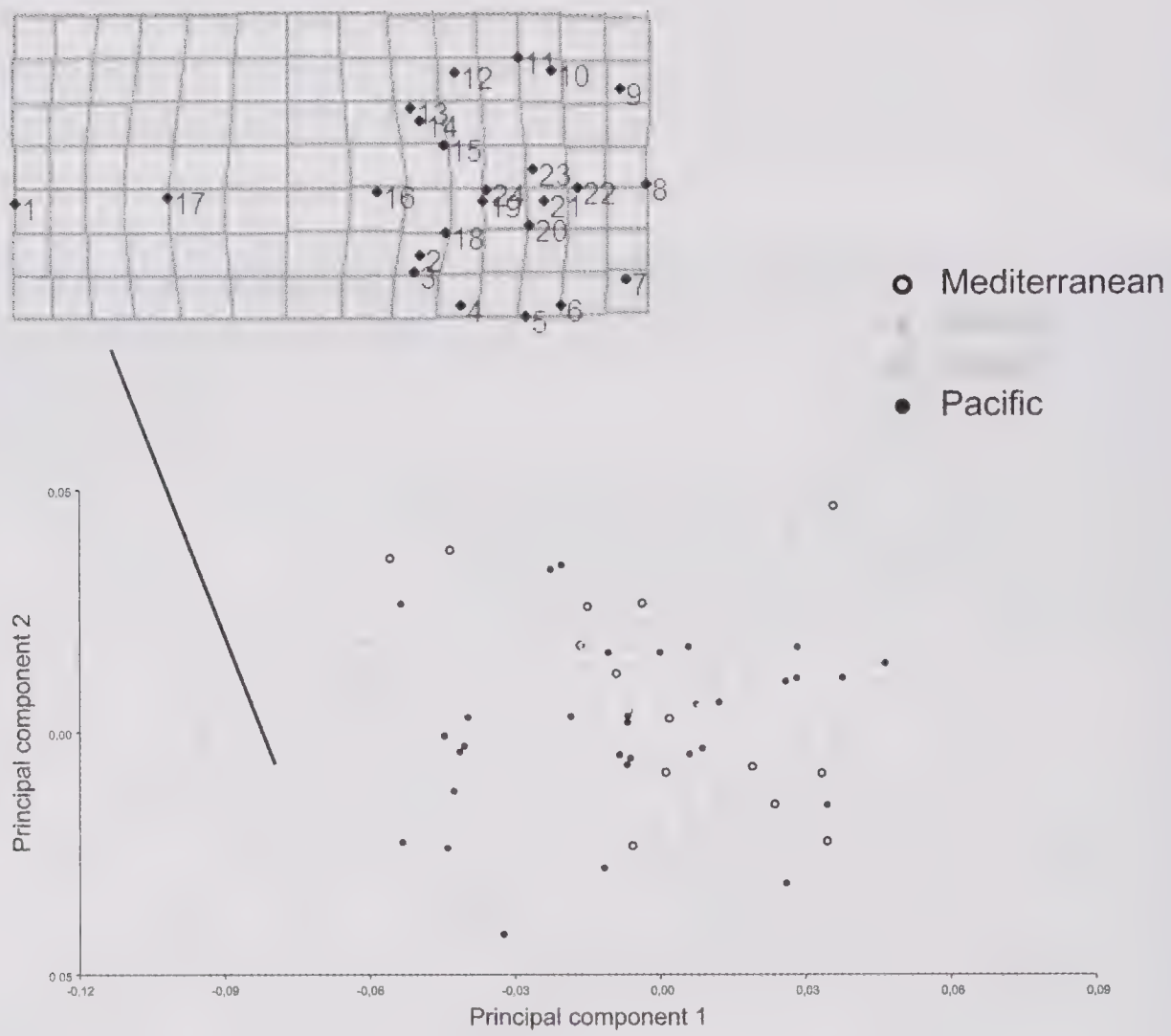


Fig. 2 – Results from PCA run on the residuals from GPA for the dorsal view of the skull: the deformation grid on the left is referred to the Indian population.

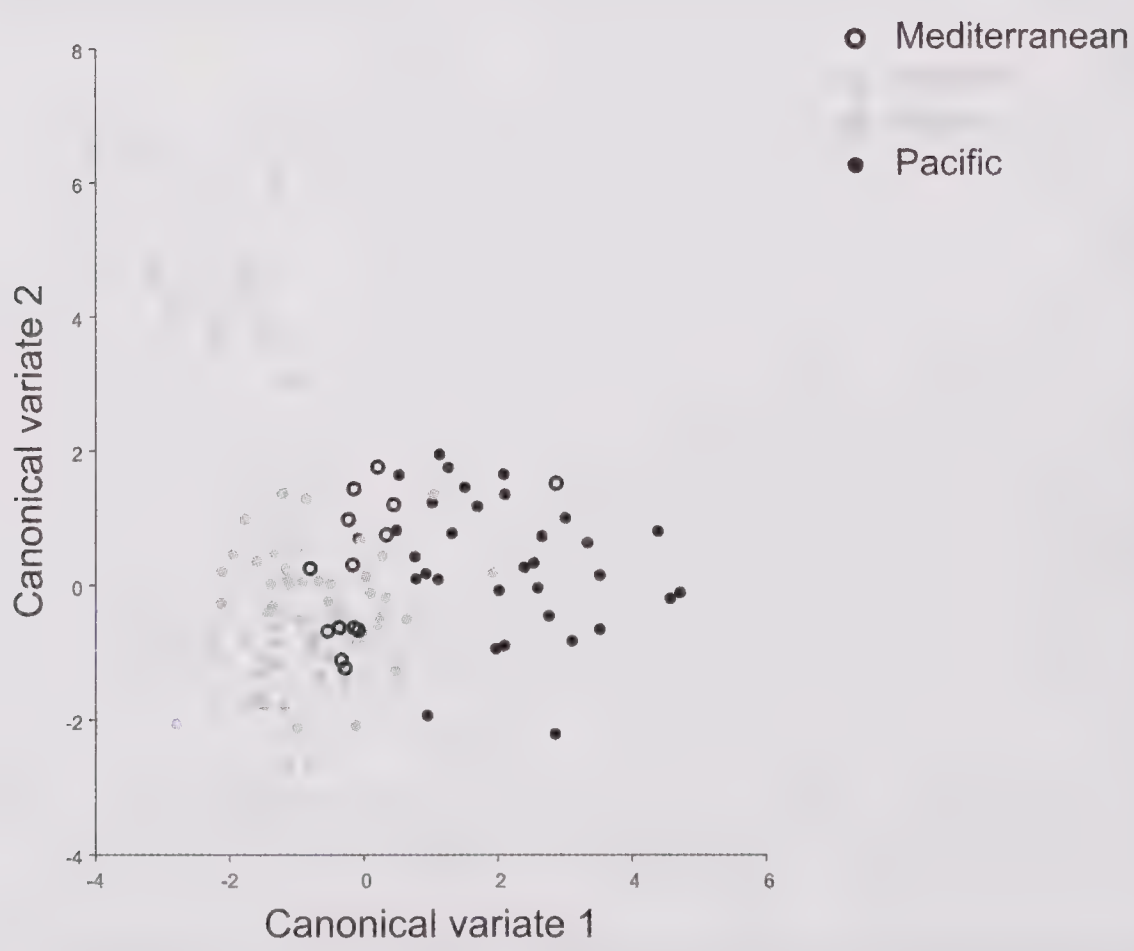


Fig. 3 – Results from the first two canonical axes extracted from residual from GPA. Symbols refer to the groups are the same for PCA and CVA.

4 Conclusion

Geometric morphometrics has shown significant differences in the shape of the skulls of *Delphinus delphis* populations from different geographical areas. Differences are particularly evident in the dolphins from the Indian ocean which appear the most divergent among all.

Other authors used the morphometric approach to identify shape differences between populations in the same dolphin species living in different geographical area "[7], [12], [13]" and also for phylogenetic and evolutionary studies "[2]". Many papers also underline the importance of morphometric analysis to support the genetic, ecological and ethological results as a powerful tool to describe and understand the mechanism of morphological differentiation "[1], [5], [8]".

These preliminary results show the need to include the other projections of the skulls to better elucidate the degree and the pattern of geographic variation, as well as adaptive traight involved in this pattern and to analyse in depth the degree and pattern of asimmetry in the region involved in the acousticmotor complex.

Acknowledgement

The authors wish to thank Dr. Graca Ramalhinho of the Natural History Museum of Lisbon, Dr. Hans J. Baagøe, Mr. Mogens Andersen and Mrs. Katrine Mohr of the Zoological Museum of Copenhagen, Dr. Ronald Vonk, Dr. Wendy van Bohemen, Dr. Roland Sluys of the Zoological Museum of Amsterdam, all the curators of the Italian Natural History Museums, Dr. Luigi Cagnolaro for information on the Italian cetological collections and Dr. Andrea Cardini for the Synthesys supporting statement and technical suggestions. This work was supported in part by a grant from the European Commission's Integrated Infrastucture Initiative programme SYNTHESYS.

References

[1] D. C. Adams, F. J. Rohlf and D.E. Slice, "Geometric morphometrics: 10 years of progress following the 'revolution'", *Ital. J. Zool.*, vol. 71, pp. 5-16, 2004.

[2] A. R. Amaral, M. M. Coelho, J. Marugàn-Lobon and F. J. Rohlf, "Cranial Shape Differentiation in Three Closely Related Delphined Cetacean Species: insights into Evolutionary History", *Zoology*, vol. 112, pp. 38-47, 2009.

[3] G. Bearzi, "*Delphinus delphis (Mediterranean subpopulation)",* IUCN Red List of Threatened Species, http://www.iucnredlist.org, 2009.

[4] G. Bearzi, R. R. Reeves, G. Notarbartolo di Sciara, E. Politi, A. Cañadas, A. Fratzis and B. Mussi, "Ecology, Status and Conservation of Short-beaked Common Dolphins (*Delphinus delphis*) in the Mediterranean Sea", *Mammal Review*, vol. 33(34), pp. 224-252, 2003.

[5] A. Cardini, D. Nagorsen, P. O'Higgins, P. D. Polly, R. W. Thorington and P. Tongiorgi, "Detecting biological uniqueness using geometric morphometrics: an example case from the Vancouver Island marmot", *Ecology, Ethology and Evolution*, DOI 10.1111/j.1439-0469.2008.00503, 2009.

[6] A. Loy, "Morphometrics and Theriology Homage to Marco Corti", *Hystrix It. J. Mamm.*, vol. 18 (2), pp. 115-136, 2007.

[7] A. Loy, A. Tamburelli, R. Carlini and D. E. Slice, "Craniometric Variation of some Mediterranean and Atlantic Populations of *Stenella caeruleoalba* (Mammalia, Delphinidae): a 3D Geometric Morphometrics Analysis", *Marine Mammal Science*, submitted for publication. (pending publication)

[8] A. Natoli, A. Cañadas, C. Vaquero, E. Politi, P. Fernandez-Navarro and A. R. Hoelzel, "Conservation Genetics of Short-beaked Common Dolphin (*Delphinus delphis*) in the

Mediterranean Sea and in the Eastern North Atlantic Ocean", *Conserv. Genet.* DOI 10.1007/s10592-007-9481-1, 2008.

[9] P. Nicolosi, A. Loy, "Morphometric Variation of Mediterranean vs Atlantic Stocks of Common Dolphin (*Delphinus delphis* Linnaeus, 1758)", *Paleontologia I Evolució, memòria especial,* vol. 3, pp. 97-99, 2009.

[10] F. J. Rohlf, *TpsDig, TpsRelw*, Department of Ecology and Evolution. State University of New York at Stony Brook. http://life.bio.sunysb.edu/morph/, 2008.

[11] F. J. Rohlf, D. E. Slice, "Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks", *Systematic Zool.*, vol. 39, pp. 40-59, 1990.

[12] A. J. Westagate, "Geographic Variation in Cranial Morphology of Short-beaked Common Dolphins (*Delphinus delphis*) from the North Atlantic", *Journal of Mammology*, vol. 88(3), pp. 678-688, 2007.

[13] K. A. Viaud-Martinez, L. R.Jr. Brownell, A. Komnenou and A. J. Bohonak, "Genetic Isolation and Morphological Divergence of Black Sea Bottlenose Dolphins", *Biological Conservation,* vol. 141, pp. 1600–1611, 2008.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 269-273.
ISBN 978-88-8303-295-0. EUT, 2010.

DNA barcoding: theoretical aspects and practical applications

Maurizio Casiraghi, Massimo Labra, Emanuele Ferri,
Andrea Galimberti, Fabrizio De Mattia

Abstract — DNA barcoding is a molecular-based identification system, recently introduced in the scientific community. The method is not completely new to science, but the real innovation is not in the discrimination system itself: DNA barcoding can be considered as the core of an integrated taxonomic system, where bioinformatics plays a key role. Time is now ripe for a real collaboration of all the different forces working in taxonomy, towards a "next generation systematics".

Index Terms — DNA barcoding, DNA taxonomy, molecular identification, species identification.

1 Introduction

The classification and monitoring of biodiversity are playing a key role in different contexts (e.g.: biological, social, economical), even if several aspects linked to these topics are far to be completely understood. A common assumption is that the central unit of taxonomy is the species, and the unequivocal association of a scientific name to a biological entity is an essential step to build a reliable reference system of biological information [1].

In the last 250 years, since Carl Linnaeus' classification system, about 1.7 million species have been formally described by taxonomists, but it is largely accepted that this number probably represents only a small fraction of the real biodiversity present on the planet (presently estimated in tens of millions of species) [2]. To help discovering this hidden biodiversity and in order to provide a useful and standardized tool for species identification, a molecular and bioinformatical tool called DNA barcoding has been proposed in 2003 [3].

The basic idea of this approach is quite simple (and not completely new

The authors are with the Department of Biotechnologies and Biosciences, University of Milan-Bicocca, Milan, Italy. E-mail:maurizio.casiraghi@unimib.it, tgalimba@gmail.com.

to science): through the analysis of the variability in a single or in a few standard molecular marker(s), it is possible to discriminate biological entities (hopefully belonging to the taxonomic rank of species). This method relies on the assumption that the genetic variation between species exceeds that within species. Consequently, the ideal DNA barcoding analysis mirrors the distributions of intra- and interspecific variabilities separated by a distance called 'DNA barcoding gap' [4], [5]. The original idea was to apply DNA barcoding systematically to all metazoans, by the use of one or a few (mitochondrial) markers (e.g. coxI, [1]). Rapidly, but with less coherent results, the idea was extended to flowering plants [6], [7] and fungi [8], and now the DNA barcoding initiative can be considered as a tool suitable for all of the tree kingdoms. Efforts in DNA barcoding development and management are coordinated by the Consortium for the Barcode of Life (CBoL; http://barcoding.si.edu/).

One of the major properties of a DNA barcode is the possibility to easily associate all life history stages and genders, to identify organisms from part/pieces, or to discriminate a matrix containing a mixture of biological species. Quite soon it became clear that DNA barcoding was suitable for two different purposes: (1) the molecular identification of already described species [1], and (2) the discovery of undescribed species [9].

A lot of rumours raised around this approach, but what is the revolution introduced by DNA barcoding? In our opinion, the big leap forward is not only the discrimination power itself, but the joint use of three innovations of modern taxonomy: (1) molecularization (i.e. the use of variability in a molecular marker as a discriminator); (2) computerization (i.e. the non redundant transposition of data using informatic supports) and (3) standardization (i.e. the extension of the approach to vast groups of not deeply related organisms). For the first time, by DNA barcoding, it is possible to introduce in taxonomy a generalization, allowing researchers specialized in different fields to work on a shared framework.

In the space of few years, DNA barcoding has moved from fantasy to reality. In some of the first enthusiastic reports, DNA barcoding was even claimed as the way to make true the dreams of Gene Roddenberry, the creator of the science fiction drama Star Trek: the creation of a tool for organism identification, the DNA barcoder, as a homologous to the fictional Tricorder [10]. A few years later we are not yet in the spaceship Enterprise, but DNA barcoding has deeply impacted the scientific community, becoming a widely used approach.

Presently, the most relevant DNA barcoding tool, The Barcode of Life Data Systems, BOLD (http://www.barcodinglife.org/, [11]) is still in constant evolution and update.

2 The 'DNA barcoding molecular entity' versus 'species' debate

Most of the questions raised by the use of DNA barcoding are directly linked to the essence of an identification method. In a strict sense, to identify simply means to differentiate, but the choice of the discriminator is essential, because the difficulty is in giving a biological meaning to what it has been discriminated [12].

Even if not always fully acknowledged, DNA barcoding implies two different

approaches to discrimination. DNA barcoding sensu stricto is a simple sorting method that could differentiate biological entities. It is not significantly different from a dichotomic key in the traditional taxonomical framework. On the other hand, DNA barcoding sensu lato represents a system that reflects the true sense of taxonomy. The discrimination method itself can be considered as an epiphenomenon - and the subject of major criticisms (DNA barcoding sensu stricto) - but it also becomes a system implementing all the aspects of taxonomy towards the representation of the living world as a whole (DNA barcoding sensu lato). It should be clear to users which kind of DNA barcoding philosophy they are going to adopt.

3 The 'DNA barcoding molecular entity' versus 'species' debate

It is well known that no identification method (morphological, biochemical, genetic, etc.) can truly identify species, because species are entities in continuous evolution and it is theoretically impossible to define statically such a dynamic matter. DNA barcoding, in its original generalization, follows the typological species approach, a concept that theoretically fails because it freezes the evolutionary continuum of species. To cope with these limitations, some development of DNA barcoding has shifted towards other species concepts [13].

The entities identified by molecular approaches have been named in several ways: 'Genospecies'; 'Phylospecies', 'Recognizable Taxonomic Units', RTUs, 'Phylotypes' sensu, 'Molecular Operational Taxonomic Units', 'MOTUs' [12]. A general naïve assumption considers 'molecular entities' and 'species' as synonyms. This is the (almost) insurmountable problem for DNA barcoding sensu stricto: the biological meaning of the identified ranks cannot be directly derived, unless we have clearly and unequivocally linked a species to the variability pattern of a single DNA barcoding marker. In all other cases, we need DNA barcoding sensu lato [12].

The identification and then the interpretation of molecular entities is the main goal of DNA barcoding. This can be reached only by users with a sound theoretical background on what this technique is able to identify.

4 The choice of the Barcode marker

DNA barcoding is not *coxI* only. A precise portion in the 5' end region of this mitochondrial gene has been proposed as a standard for metazoans [1]. Even if *coxI* has proven to be useful to discriminate species in most of the tested groups, its limits in some animal taxa are already evident (e.g. [14]). The choice of regions usable for DNA barcoding has been little investigated in many other eukaryotes. For instance, a marker was already available in fungi: the nuclear ITS region, which has been now confirmed as the main DNA barcode for this group [15]. In terrestrial plants, compared to animals, mitochondrial DNA has slower substitution rates and shows intramolecular recombination [16]. The search for an analogous to *coxI* or ITS in plants, that matches with the DNA barcoding criteria, has focused attention on the plastid genome. Several plastid genes have been proposed, such as the most conserved *rpoB, rpoC1*

and *rbcL* or a section of *matK* showing a rapid rate of evolution, but in some plant families these genes showed amplification problems. At the same time, the intergenic spacers such as *trnH-psbA, atpF-atpH* and *psbK-psbI* were also tested for their rapid evolution [17. Recently, the CBoL Plant Working Group [18] provided a recommendation on a standard plant barcode suggesting the 2-locus combination of *rbcL* and *matK*.

5 Biological and bioinformatical Repositories for DNA barcoding data

DNA barcoding data are meant to be easily and widely accessed. To reach this aim, a proper sequence submission procedure is available for GenBank (http://www.ncbi.nlm.nih.gov/WebSub/?tool=barcode). This procedure slightly modifies the standard sequence submission procedure, introducing a DNA barcoding label to the sequence in order to simplify database querying and searching. Moreover, additional data are requested to link barcode sequence data to its voucher specimen. This standardization is mirrored by the establishment of the Registry of Biological Repositories initiative (http://www.biorepositories.org/), an on-line registry of organisms linked to DNA sequences. DNA barcoding sequences can also be deposited as projects in BOLD databases, characterized by an automatic submission tool to publish sequences to GenBank. By December 2009, BOLD database encompassed more than 760,000 sequences, corresponding to more than 65,300 formally described 'species'. The amount of data managed by the BOLD database is impressive: it collects, for a large amount of deposited barcode sequences, specimen details such as morphology, photographs, geographical distribution, collection points and others [11].

6 Conclusion

DNA barcoding is not a "perfect" method, but it has deeply impacted the scientific community, becoming a widely used approach, characterized by many relevant aspects of uniformity and generalization. A critical knowledge of the method is essential for a proper use of it.

Acknowledgement

The authors are grateful to the ZooPlantLab staff, students and supporters.

References

[1] Q. D. Wheeler, "Taxonomic triage and the poverty of phylogeny". *Phil. Trans. R. Soc. Lond.* B, vol. 359, pp. 571-583, 2004.

[2] R. Vernooy, E. Haribabu, M. R. Muller, J. H. Vogel, P. D. N. Hebert, et al., "Barcoding Life to Conserve Biological Diversity: Beyond the Taxonomic Imperative". *PLoS Biol.*, vol. 8(7): e1000417. doi:10.1371/journal.pbio.1000417, 2010.

[3] P. D. N. Hebert, A. Cywinska, S. L. Ball, et al., "Biological identifications through DNA barcodes". *Proc. R. Soc. London, Biol. Sci. Series* B, vol. 270, pp. 313–321, 2003.

[4] C. P. Meyer and G. Paulay, "DNA Barcoding: error rates based on comprehensive sampling". *PLoS Biol.*, vol. 3, e422, 2005.
[5] M. Wiemers and K. Fiedler, "Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae)". *Frontiers in Zoology*, vol. 4, p. 8, 2005.
[6] W. J. Kress, K. J. Wurdack K., E. A. Zimmer, et al., "Use of DNA barcodes to identify flowering plants". *PNAS*, vol. 102, pp. 8369-8374, 2005.
[7] M. L. Hollingsworth, A. Clark, L. L. Forrest, et al., "Selecting barcoding loci for plants: evaluation of seven candidate loci with species level sampling in three divergent groups of land plants". *Mol. Ecol. Res.*, vol. 9, pp. 439-457, 2009.
[8] X. J. Min and D. A. Hickey, "Assessing the effect of varying sequence length on DNA barcoding of fungi". *Mol. Ecol. Notes.,* vol. 7, pp. 365–373, 2007.
[9] P. D. N. Hebert, E. H. Penton, J. M. Burns, et al., "Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*." *PNAS*, vol. 101, pp. 14812-14817, 2004.
[10] K. J. Gaston and M. A. O'Neill, "Automated species identification: why not?" *Phil. Trans. R. Soc. Lond.* B, vol. 359, pp. 655-667, 2004.
[11] S. Ratnasingham and P. D. N. Hebert, "BOLD: The Barcode of Life Datasystem (www.barcodinglife.org)". *Mol. Ecol. Notes*, vol. 7, pp. 355-364, 2007.
[12] M. Casiraghi, M. Labra, E. Ferri, A. Galimberti and F. De Mattia, "DNA barcoding: a six-question tour to improve users' awareness about the method." *Brief Bioinform.*, vol. 11(4), pp. 440-453. Epub 2010 Feb 15, 2010.
[13] J. M. Padial, A. Miralles, I. De la Riva and M. Vences, "The integrative future of taxonomy." *Front Zool.*, vol. 7, p. 16, 2010.
[14] T. L. Shearer and M. A. Coffroth, "Barcoding corals: limited by interspecific divergence, not intraspecific variation." *Mol. Ecol. Resour.*, vol. 8, pp. 247-255, 2008.
[15] R. H. Nilsson, M. Ryberg, E. Kristiansson, et al., "Taxonomic Reliability of DNA Sequences in Public Sequence Databases: A Fungal Perspective." *PLoS One*, vol. 1, e59, 2006.
[16] G. D. D. Hurst and F. M. Jiggins, "Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts." *P. Roy. Soc. Lond.* B Bio, vol. 272, pp. 1525-1534, 2005.
[17] A. J. Fazekas, K.S. Burgess, P. R. Kesanakurti, et al. "Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well." *PLoS One*, vol. 3, e2802, 2008.
[18] CBoL Plant Working Group. "A DNA barcode for land plants." *PNAS*, vol.106, pp. 12794-12797, 2009.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 275-280.
ISBN 978-88-8303-295-0. EUT, 2010.

Strength and Limitations of DNA Barcode under the Multidimensional Species Perspective

Valerio Sbordoni

Abstract — DNA barcoding aims at providing an efficient method for species-level separation using a partial sequence of the mitochondrial COI gene. The efficiency of the barcode in separating species is based on the amount of genetic distance among samples. While in many taxa the species can be efficiently identified through the barcode, other situations cannot be treated by this approach. The causes for such discrepancy appear to be mostly related to the nature of speciation events and to the different roles of the genetic system, natural selection and evolutionary time. Thus, DNA barcode represent just one important descriptor in the framework of the multidimensional species approach.

Index Terms — DNA barcode, molecular systematics, species concepts, taxonomic procedures.

1 Introduction

Since 2003, many research groups started to accumulate molecular data with the aim of setting up a sort of inventory of life that might itemize biodiversity as a sequence of species-specific DNA. In particular, Paul Hebert from the Canadian University of Guelph [1], [2],proposed to use a sequence of the COI mitochondrial gene, codifying for cytochrome oxidase 1,as a "molecular signature" to identify a species. The selection of this gene for exploring limits between species allows for the practical advantages of using mitochondrial DNA together with the previous wide use of this gene in a large variety of organisms. COI sequences are also currently used at different taxonomic levels, in phylogeny, phylogeography and population genetics studies, due to the great advantage offered by the availability of amplification protocols, as well as a large number of sequences ready for barcoding.

At present there are many ongoing DNA barcoding projects reported on the

V. Sbordoni is with the Department of Biology, University of Roma Tor Vergata, Italy.
E-mail: valerio.sbordoni@uniroma2.it.

website of the Barcode of Life Data Systems (www.boldsystems.org), an online workbench that supports collection, management, analysis, and use of DNA barcodes. An enormous bulk of barcode data for a wide array of organisms has already been made available to the scientific community. Thus, a sequence of the mitochondrial COI gene has become the most used mitochondrial marker, especially for animals. The same marker, preferably associated with nuclear DNA sequences, is commonly used also in more wide phylogenetic studies. The choice of a marker specific for plants and fungi is more problematic, but it currently seems oriented, at least for angiosperms, to the *trnH-psbA* sequence, an intergenic spacer of plastidic DNA [3], [4].

2 Molecular Systematics and the DNA Barcode

By identifying genetic differentiation threshold values that would include individuals of the same species, the barcode approach allows investigating and analysing in some detail inter-species delimitation and its many related problems. A great advantage offered by DNA barcoding is the possibility of identifying cryptic species, that is to distinguish as belonging to different species individuals that, due to their similar morphology, were considered as belonging to a unique species. This would be feasible by identifying genetic distance limit values within which two individuals can be considered as belonging to the same species, while outside these limits they should be considered as belonging to different species. Although the limit seems to be somehow taxon dependent, it has been observed that a value of genetic distance between two DNA-barcode sequences equal to or higher than three per cent ($D \geq 0.03$) identifies distinct species.

However, it has become clear that the currently used genetic distance approaches by means of DNA barcodes have strong limitations, particularly when it comes to defining species boundaries [5]. One reason is that mtDNA rates of evolution vary substantially between and within species and between different groups of species, thus resulting in broad overlaps of intra- and interspecific distances [6], [7], [8]. But there are other reasons worth to be briefly discussed.

To what extent DNA barcode may solve the problem of identification of interspecific limits? And, will this approach be applicable to the different typologies of species? Based on the data produced up to day, it seems that DNA barcoding may solve problems of specific classification in a wide range of organisms and situations, but many taxa cannot be treated by this approach [9].

Let's try to understand why. Biologists and evolutionists know that species are the result of historical processes, mainly speciation and range dynamics. The knowledge of these processes is of help for inferring the essence of species whose properties reflect the "signature" of speciation where the following factors have more or less predominant roles:

1. The genetic system.
2. Natural selection.
3. Evolutionary time.

The genetic system, *i.e.* the organization of genes and of other chromosomal

genetic material may favour, by acceleration, the speciation process especially when it is based on Robertsonian translocations, or inversion polymorphisms or, as with the extreme case of instantaneous speciation, on allopolyploid mechanisms, so frequent in plants.

Natural selection may have a determinant role in those sympatric speciation phenomena connected to the shift of the trophic niche, that are frequent especially in phytophagous insects. In these cases selective pressure may significantly accelerate the adaptive diversification process of genotypes in their new evolutionary path.

Finally, in the absence of diverging selective pressure, even the simple allopatric condition of two originally conspecific populations may lead to speciation in the long run if the genic flow barrier persists. This is where the evolutionary time finds its role. The longer the isolation time, the higher the number of different nucleotidic substitutions in the DNA sequences of the two genic pool involved. Accumulation of diverging mutations in the whole genome will inevitably lead to speciation by genetic drift, and the speed of this process will be inversely proportional to the population effective size.

This brief analysis of times and modes of speciation is of help for interpreting the biological meaning of genetic distance estimates based on the DNA barcode concept. A limit of 3% genetic divergence is relatively well working to separate species that are the result of geographic speciation events driven by gradual accumulation of diverging mutations. Emblematic examples are diverging populations and species of animals adapted to cave life [10], with special regard to the thoroughly studied *Dolichopoda*cave crickets [11], [12], [13].

On the other hand, much smaller values, up to a ten factor, do not allow the discrimination by barcoding of species differing by one or more chromosomal inversions, as it is the case with some *Anopheles* [14], or of recent species originated sympatrically by adaptive shift, as it has been reported in fruit flies of the genus *Rhagoletis* [15]. Most probably, in the case of *Dolichopoda*, the divergence process may have involved most of the genome and, consequently, mitochondrial genes as well. Conversely, in the other two examples, the divergence should have involved, in relatively short times, only a few genes target of selection, or particular combinations of these genes, showing no effects in mitochondrial DNA sequences. Indeed, mitochondrial DNA is often used as a molecular clock [16], [17].

3 The Multidimensional Approach

The debate on species' properties shows many different aspects: a more classical one opposing the "biological concept of species, BSC" to the "typological, TSC" one [18], and a more modern one that seeks a conflict between BSC and the "phylogenetic concept" [19].

But when analysing the nature of these contrasts it must be agreed that the three concepts focus on properties of the species that do not contrast and which are coherent with a vision of the species as the *in itinere* product of the evolutionary process. Therefore it should be agreed that a species, beside representing *"the smallest monophyletic group of common ancestry"*

[20], expresses its distinct gene pool (BSC) by having eventually acquired the distinctive characters emphasized in the typological species concept.

Actually, in routine work, many taxonomists tend to use operational approaches toward species, based on morphological characters that are unique and shared or their hierarchies. Although this is not explicitly declared, these species are based on philosophies ranging from the classic typological concept up to the phylogenetic and the phenetic ones. Traditional phenetic definitions of species are based essentially on the numeric recognition of intervals separating clusters of phenetically similar individuals [21], [22]. For this evaluation many different, "not weighed" types of taxonomic characters are examined, but their biological meaning is not evaluated.

The phenetic "concept" has been strongly criticized because it was not considered as sufficient to describe the complex interrelations existing among clusters of similar populations. Nonetheless, this kind of approach has operational advantages of some practical value and consequently it is widely adopted and used by systematists.

Being an intrinsically complex entity, the species requires a multidimensional approach taking into account the whole set of taxonomic characters. Nowadays, the huge progress in multivariate analysis, together with the wide choice of technologies to measure parameters related to the ecology of niche, or to sexual behaviour, and/or to many other crucial features of species-specificity, have richly endowed the kit of characters available to the modern taxonomist. The technical and conceptual progress in the field of molecular biology has made relatively easy and rapid both the acquisition and the phylogenetic interpretation of sequence data containing an enormous quantity of information. The routine use of these characters has substantially empowered the understanding of the species' genetic structure and has brought to the discovery of the existence of cryptic species.

However, as previously discussed, DNA sequences are not the only depositories of evolutionary history. Any other kind of character is potentially suitable to give its own contribution to the species definition. For instance, much of the ecological role (niche) of an organism, is written in its morphology, although there is no assurance that the ecological divergence between two similar species corresponds to a genetic gap, even though this coincidence shows up in the majority of cases. The diagnostic value of each character varies by taxon or specific evolutionary, geographical and ecological situations. For most taxa specialists know very well which characters are more kin to represent the species' biological properties.

This discussion makes us re-consider, although with due caution and adjustments, the usefulness of the phenetic approach. Since different descriptors have different values for a given organism, one cannot rely on automatic discriminating procedures, while one can rely on the availability of the whole set of algorithms and multivariate procedures set up in the systematic and ecological fields. Yet, the responsibility of the final decision will inevitably have to rely on the competence and experience of the specialist. It must be stressed that different taxonomic characters do not necessarily vary in a coordinated way, yet they are often conflicting. Both evaluation and weighing of characters

have always had and continue to have a fundamental role in the process of species' delimitation [11].

Based on this logic and premises, species are considered as "clusters of individuals that are effectively separated from other clusters in the space defined by their descriptors" [23]. Alike the preceding phenetic definitions, species are seen as clouds of probability in an hyperspace. Here though characters are weighed and a value is assigned to genetic, and inter-reproductive descriptors, *i.e.* exactly those characterising the species as a monophyletic cluster, as a cluster of genotypes and as a cluster of individuals sharing a special relation with their environment. An "ad hoc" reduction of this hyperspace makes this multidimensional concept operative. For instance, the typical biological species becomes a particular case where intra-population genetic and reproductive relationships are quantified and analysed as a sub-set of a wider set of descriptors. Yet, the use of the multidimensional approach should be particularly useful for organisms with asexual or uniparental reproduction, including bacteria, protists, fungi, rotifera and many parthenogenetic taxa, to which it is traditionally difficult or impossible to apply the biological concept of species, thus overcoming the tie of amphigonic reproduction and allowing, not only in theory, the evaluation of clusters defined by appropriate descriptors. The literature on taxonomy, and not only the recent one, offers many examples of this approach, adopted with success in cave crickets [11], butterflies [24], fishes [25], fossil Ostracoda [26], Rotifera[27], etc.

Many species' definitions, privileging either properties, can be accommodated within this approach, but I want here to recall in particular a somewhat unknown definition by Alfred Russell Wallace, incidentally quoted in one of his writings where he disputes with Galton: "*A species … is a group of living organisms, separated from all other such groups by a set of distinctive characteristics, having relations to the environment not identical with those of any other group of organisms, and having the power of continuously reproducing its like*" [28].

This definition, dated many years before the Synthetic Theory, refers to all the emerging properties of species: a set of distinctive characters (highlighted by the typological concept), the relationship with the environment (ecological concept), and finally the power of reproducing its own characteristics, which implies the properties of the hereditary material. Compared with many others, Wallace's definition has certainly the merit of stressing multidimensionality, a concept that expresses the best operational solution to the problem of the delimitation of species.

References

[1] P. D. N. Hebert, A. Cywinska, S. L. Ball and J. R. DeWaard, "Biological Identifications Through DNA Barcodes", *Proc. R. Soc. Lond. B.*, vol. 270, pp. 313-321, 2003.

[2] S. Ratnasingham and P. D. N. Hebert, "Barcoding Bold: The Barcode of Life Data System", *Molecular Ecology Notes*, doi: 10.1111/j.1471-8286.2006.01678.x, 2007.

[3] W. J. Kress, K. J. Wurdack, E. A. Zimmer, L. A. Weigt and D. H. Janzen, "Use of DNA Barcodes to Identify Flowering Plants", *PNAS*, vol. 102, pp. 8369-8374, 2005.

[4] M. W. Chase and M. F. Fay, "Barcoding of Plants and Fungi", *Science*, vol. 325, pp. 682–683, 2009.

[5] J. D. Witt, D. L. Threloff and P. D. Hebert, "DNA Barcoding Reveals Extraordinary Cryptic Diversity in an Amphipod Genus: Implications for Desert Spring Conservation", *Mol. Ecol.*, vol. 15, pp. 3073–3082, 2006.
[6] W. W. Kipling and D. Rubinoff, "Myth of the Molecule: DNA Barcodes for Species Cannot Replace Morphology for Identification and Classification", *Cladistics*, vol. 20, pp. 47–55, 2004.
[7] D. Rubinoff, "Utility of Mitochondrial DNA Barcodes in Species Conservation", *Conserv. Biol.*, vol. 20, pp. 1026–1033, 2006.
[8] D. Rubinoff, S. Cameron and K. Will, "A Genomic Perspective on the Shortcomings of Mitochondrial DNA for "Barcoding" Identification", *J. Hered.*, vol. 97, pp. 581–594, 2006.
[9] J. Waugh, "DNA Barcoding in Animal Species: Progress, Potential and Pitfalls", *BioEssays*, vol. 29, pp. 188–197, 2007.
[10] V. Sbordoni, "Advances in Speciation of Cave Animals". In: C.Barigozzi (ed.), *Mechanisms of Speciation*, New York, A.R. Liss Inc., pp. 219-240, 1982.
[11] V. Sbordoni, G. Allegrucci and D. Cesaroni, "A Multidimensional Approach to the Evolution and Systematics of *Dolichopoda* Cave Crickets". In: G. M. Hewitt et al. (eds.), *Molecular Techniques in Taxonomy,* NATO ASI Series, vol. H57, Berlin, Springer, pp. 171-199, 1991.
[12] G. Allegrucci, M. Rampini, P. Gratton, V. Todisco and V. Sbordoni, "Testing Phylogenetic Hypotheses for Reconstructing the Evolutionary History of *Dolichopoda C*ave Crickets in the Eastern Mediterranean", *J. Biogeography*, vol. 36, pp. 1785-1797, 2009.
[13] L. Martinsen, F. Venanzetti and A. Johnsen, V. Sbordoni and L. Bachmann "Molecular Evolution of the pDo500 Satellite DNA Family in Dolichopoda Cave Crickets (Rhaphidophoridae)", *BMC Evolutionary Biology*, vol. 9, 301 (14 pp.), 2009.
[14] M. Coluzzi, A. Sabatini, A. della Torre, M.A. Di Deco and V. Petrarca, "A Polytene Chromosome Analysis of the *Anopheles gambiae* Species Complex", *Science*, vol. 298, pp. 1415–1418, 2002.
[15] J. J Smith and G.L. Bush, "Phylogeny of the *Genus Rhagoletis* (Diptera: Tephritidae) Ifrom DNA Sequences of Mitochondrial Cytochrome Oxidase II", *Mol. Phyl. Evol.*, pp. 33-43, 1997.
[16] L. Bromha and D. Penny, "The Modern Molecular Clock", *Nature Rev. Genet.*, vol. 4, pp. 216–224, 2003.
[17] A. Caccon and V. Sbordoni, "Molecular Biogeography of Cave Life: a Study Using Mitochondrial DNA from Bathysciine Beetles", *Evolution*, vol. 55, pp. 122–130, 2001.
[18] E. Mayr, *Systematics and the Origin of Species, from the Viewpoint of a Zoologist,* New York, Columbia University Press, 1942.
[19] M. J. Donoghue, "A Critique of the Biological Species Concept and Recommendations for a Phylogenetic Alternative", *Bryologist*, vol. 88, pp. 172-181, 1985.
[20] K. de Queiro and M. J. Donoghue, "Phylogenetic Systematics and Species Revisited", *Cladistic*, vol. 6, pp. 83-90, 1990.
[21] C. D. Michener, "Diverse Aapproaches to Systematics", *Evol. Biol.*, vol. 4, pp. 1-38, 1970.
[22] E. H. A. Sneat and R. R. Sokal, *Numerical Taxonomy,* W. H. Freeman, San Francisco, 1973.
[23] V. Sbordoni, "Molecular Systematics and the Multidimensional Concept of Species", *Biochemical Systematics and Ecology*, vol. 21, pp. 39-42, 1993.
[24] D. Cesaroni, M. Lucarelli, P. Allori, F. Russ and V. Sbordoni, "Patterns of Evolution and Multidimensional Systematics in Graylings (Lepidoptera: *Hipparchia*)", *Biol.J.Linn.Soc.*, vol. 52, pp. 101-119, 1994.
[25] M. Barluenga, K. N. Stolting, W. Salzburger, M. Muschic and A. Meyer, "Sympatric Speciation in Nicaraguan Crater Lake Cichlid Fish", *Nature*, vol. 439, pp. 719-23, 2006.
[26] M. Gross, K. Minati, D.L. Danielopo and W. E. Piller, "Environmental Changes and Diversification of *Cyprideis* in the Late Miocene of the Styrian Basin (Lake Pannon, Austria)", *Palaeobiodiversity and Palaeoenvironments*, vol. 88, pp. 161-181, 2008.
[27] D. Fontaneto, E. A. Herniou, C. Boschetti, M. Caprioli, G. Melone, C. Ricc and T. G. Barraclough, "Independently Evolving Species in Asexual Bdelloid Rotifers", *Plos Biology*, doi: 10.1371/journal.pbio.0050087, 2007.
[28] A. R. Wallace, "The Method for Organic Evolution", *Fortnightly Review (N.S.)*, vol. 57, pp. 435-445, London, 1895.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 281-287.
ISBN 978-88-8303-295-0. EUT, 2010.

DNA Barcoding and Phylogeny of Patellids from Asturias (Northern Spain)

Yaisel Juan Borrell, Fernando Romano, Emilia Vázquez, Gloria Blanco, Jose Antonio Sánchez Prado

Abstract — The main role for genetics in marine invertebrates is the identification of species and groups of interbreeding individuals. In Asturias exists an ancient culinary tradition of consumption for limpets (*Patella s.s.*) but there is a lack of studies about these species. We have sampled Asturian marine *Patella s.s.* specimens and conducted sequencing of the mtDNA COI gene. We have confirmed the presence of four *Patella s.s.* in the Asturian coasts (*P. vulgata*, *P. depressa*, *P.aspera*, and *P. rustica*). Our work raises concerns about the current state of the *P. vulgata* populations in Asturias, where it is exploited, due to its low levels of genetic variation. Our phylogenetic analyses using Bayesian approaches confirmed that patellids belong to four different clades but gives a new version about how these clades are related inside the genus aiming for the necessity of more work to address this issue.

Index Terms — Patellids, COI gene, Bayes phylogenies, Asturias.

1 Introduction

The family *Patellidae* contains most of the common limpets on the temperate rocky shores of Europe. It contains around 37 species distributed in four morphological clades: *Helcion*, *Cymbula*, *Scutellastra* and *Patella* [1]. The phylogeny and taxonomy of this family has been revised and modified many times and it is not completely resolved yet [1], [2], [3]. Several morphological characters have been used to differentiate these species (e.g. radular morphology, headfoot, sperm, etc); however, the external form of the shell had been the principal character used in the species-level taxonomy of patellids [1], [4]. Despite this, it is known that shells are highly variable and usually lead to taxonomy confusions (see Mauro et al. [4]).

The main role for genetics in marine invertebrates is the identification of species and groups of interbreeding individuals [5]. A public library of sequences linked to named species, and the promotion of portable devices for DNA barcoding, will

All authors are with the Laboratorio de Genética Acuícola, Departamento de Biología Funcional, IUBA, Universidad de Oviedo, 33071 Oviedo, Spain. E-mail: borrellyaisel@uniovi.es.

also considerably help in the management and conservation of marine species [6]. The cytocrome c oxidase I gene (COI) seems to be a useful genetic tool to be used in patellids molecular genetics. The universal primers for this gene are very robust and COI appears to possess a greater range of phylogenetic signal than any other mitochondrial gene [7], [8]. Kouffopanou et al. [2] used 12S and 16S genes for constructing the first molecular phylogeny of patellids. More recently Sá-Pinto et al. [3] used also COI and pointed to *Patella s.s.* showing five strongly supported clades (I: *P. candei*, *P. lugubris, P. caerulea, P. depressa*; II: *P. ulyssiponensis* (*P. aspera*); III: *P. vulgata*. IV: *P. rustica, P. ferruginea*; V: *P. pellucida*).

The *Patella s.s.* show problems for their conservation and management today. Conservation of declining stocks of *P. candei* and *P. aspera* has become a concern on the Atlantic islands [9], while in the Mediterranean, *P. ferruginea* is seriously endangered [10] and in the eastern Pacific *P. mexicana* may be locally extinct on parts of the mainland coast of Mexico [1]. In Asturias, northern Spain, exists an ancient culinary tradition of consumption for limpets (*Patella s.s.*). Even when they have been under commercial exploitation for decades, there is not any previous genetic data about them. There is only a few data about the composition of the genus and the more recent morphological studies dates from 1980 [11], [12], [13]. It had been reported four species (*P. vulgata*, *P. depressa*, *P. aspera* and *P. rustica*) of which the principally harvested is *P. vulgata*. We have sampled Asturian marine *Patella s.s.* specimens and conducted sequencing of the mtDNA COI gene. Our aims were to identify/confirm the patellids species present in our coasts using molecular methods and to analyze the phylogenetic status of the genus using a Bayesian approach.

2 Material and methods

Fourty five *Patella s.s.* specimens were collected in four areas (western Asturias: Punta La Cruz (10 inds), central Asturias: Moniello (8 inds), Antromero (18 inds) and eastern Asturias: Tereñes (9 inds)) (Fig. 1). Patellid individuals were morfologically identified by an experienced zoologist. DNA extractions were carried out using the ZR Genomic DNA Kit™. We used the universal primers described by Folmer et al. [7] for the amplification of a 658 bp fragment of the Citochrome Oxidase subunit I gene (COI). Sequences were aligned in a multiple alignment with the BioEdit software package. After alignment and trimming, the final sequence length used was 642 bp. We used Collapse 1.2 for collapsing sequences to haplotypes. A Global search in the Barcoding of Life system (CBOL) resource for the identification of species using the COI sequences was carried out. Nucleotide (π) and haplotype (h) diversities were calculated using Arlequin 3.11. For Bayesian phylogenetic inferences MrBayes 3.1.2 was used evaluating tree topologies and models of nucleotide substitution for the aligned dataset. The model of nucleotide evolution was the general time-reversible model (GTR+G) with gamma-distributed rate variation across sites but not proportion of invariable sites.

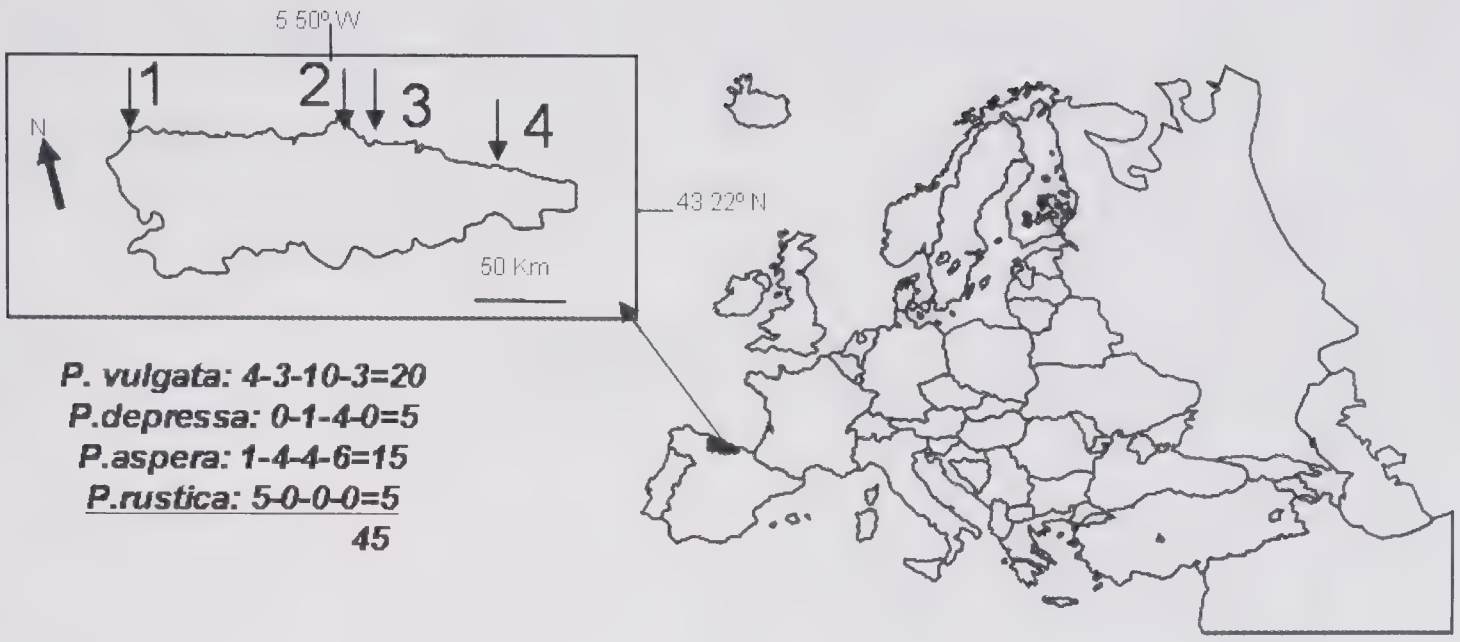


Fig. 1 – Map showing the geographic localization for patellids samples in Asturian coastal area (northern Spain) 1- Punta la Cruz. 2- Moniello. 3- Antromero 4- Tereñes. Limpet individuals assigned to each species of the *Patella* genus after *COI* identification are indicated by areas in the same order.

3 Results

The COI sequence data have been deposited in the GenBank nucleotide sequence database with accession numbers EF462952 to EF462975 (23 patellids haplotypes). The 45 limpets (*Patella s.s.)* collected in the Asturian coasts were morphologically assigned to the species *P. vulgata* (20), *P. depressa* (5), *P. aspera* (15) and *P. rustica* (5). This was concordant with assignments using the molecular method and the CBOL resource. However, two individuals showed more than a 2% of sequence divergence with its species sequences (PA-1205 and PA-3105 assigned to *P. aspera* with a 94.4% and 97.8% of similarity, respectively).

The levels of COI genetic variation for each of the mentioned above species are shown in Tab. 1. Inside the *Patella* genus we found an 8.8% and 88.1% of nucleotide (π) and haplotype (h) diversity, respectively. The most variable species was *P. aspera* (π =1.60%) while *P. vulgata* appeared as the less variable one (π =0.07%) (Tab. 1).

The COI phylogenetic tree using the Bayesian approach showed two main nodes: one contains the *P. aspera* individuals and another contains the *P. vulgata* ones and also a group of *P. depressa* and *P. rustica* individuals (Fig. 2). One individual (PA-1205), morphologically classified as *P. aspera*, was located inside the *P. vulgata* branch although it showed only a 93.6% of similarity with the available *P. vulgata* sequences (CBOL).

4 Discussion

Our sampling and the identification methods used here revealed four *Patella* species in Asturias (*P. vulgata*, *P. depressa*, *P.aspera*, and *P. rustica*) confirming

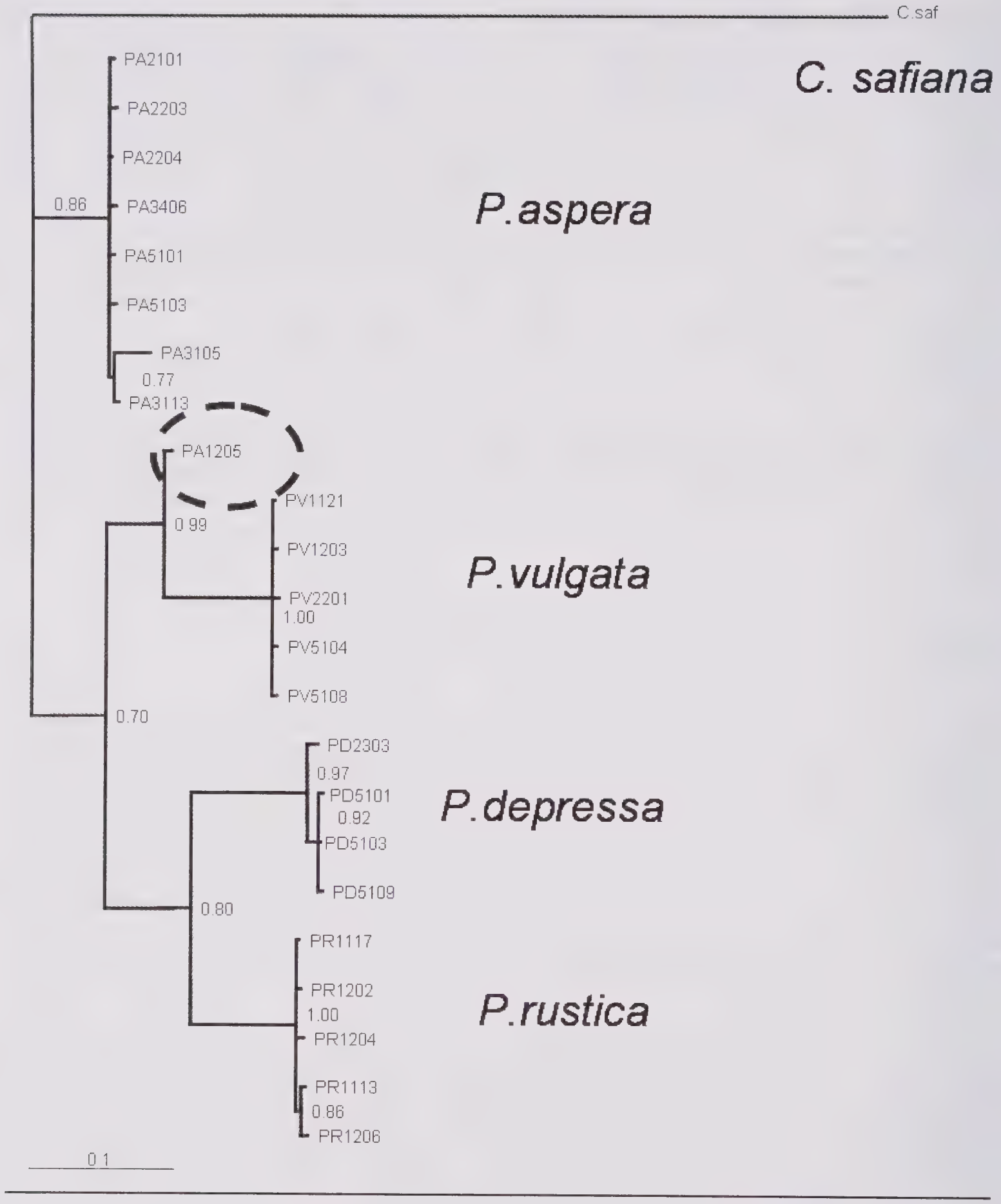


Fig. 2 – Consense Phylogenetic tree after Bayesian analysis using Cytochrome Oxidase I (*COI*) sequences in the genus *Patella*. C.*safiana* as output group. Numbers represent more than a 70% of branch support. A *P.aspera* individual grouping with the *P.vulgata* group is indicated by a discontinuous circle.

Patella Species	Gene	bp	n	Haplotypes numbers	Genbank AN	Polymorphic. sites	Ts	Tv	Sb	Id	π	h
	COI	642	45	23	EF462952-75	156	139	38	177	0	0.0885	0.8808
P.vulgata			20	5	EF462957-62	5	4	1	5	0	0.0007	0.4474
P.depressa			5	4	EF462972-75	8	5	3	8	0	0.0049	0.9000
P.rustica			5	5	EF462952-56	7	5	2	7	0	0.0046	1.0000
P.aspera			15	9	EF462963-71	68	53	16	69	0	0.0160	0.8857

Tab. 1 – Genetic variation in patellids from the Astrurian Coasts.
bp: base pairs. n: number of samples. Genbank AN: Genbank accession numbers for haplotypes. Ts: transitions. Tv: transversions. Sb: substitutions. Id: indels. π: Nucleotide diversity. h: Haplotype diversity. COI: Cytocrome Oxidase I.

previous morphological studies from the 80s [12], [13]. Two individuals fall apart from the criterion of less than a 2% of divergence for correct species classifications (following the CBOL recommendations) and the PA-1205 haplotype is clearly out of its putative species branch in the phylogenetic COI-trees. This could point to cases of genetic introgression and to the necessity to clarified and revise the taxonomic classifications in *Patella* species [3].

The patellids found showed a dissimilar species distribution in the Asturian coastal area with *P. rustica* being only present in the Asturian/Galician frontier. Possibly, this species is the most sensible to changes in the sea surface temperature that determine its reproductive success and hence its dispersal potential [14]. The commercially exploited species (*P. vulgata*) is the genetically

less variable. This raises concerns about the health of this species in Asturias.

The *Patella* genus is monophyletic inside the *Patellidae* family. Sá-Pinto et al. [3] showed five strongly supported clades for patellids although relationships between them were not well supported. Working with four species included in the Sá-Pinto et al. [3] study we have recovery the expected four clades phylogenetic tree. However, we have well supported branches (more than 80%) indicating different relationships among these clades. Our results revealed two main nodes: one included *P. aspera* (*P. ulyssiponensis*) as an entity and another included *P. vulgata, P. depressa* and *P. rustica* species. Our results differ from the closeness between *P. depressa* and *P. vulgata* species proposed by Koufopanou et al. [2] (*P. depressa* is more close to *P. rustica* in our work) and also from the relationships between clades showed by Sá-Pinto et al [3]. They showed *P. depressa* (clade I) together with *P. aspera* (Clade II) and these two grouped to *P. vulgata* (Clade III), while *P. rustica* (Clade IV) was an independent entity. It will be necessary much more work and different approaches to ascertain which the relationship between clades inside the genus *Patella* is. This will help to clarify taxonomy and will give us clues about speciation patterns and origins inside the genus. All this information will be vital for its adequate conservation and management.

5 Conclusion

Molecular methods are useful tools for species identification when morphological analyses lead to taxonomic confusions. Using COI gene sequences we have confirmed the presence of four patellids in the Asturian coasts (*P. vulgata*, *P. depressa*, *P. aspera*, and *P. rustica*). Our work raises concerns about the current state of the *P. vulgata* populations in Asturias, where it is exploited, due to its low levels of genetic variation. Our phylogenetic analyses confirmed that patellids belong to four different clades; however our work gives a new version about how these clades are related inside the genus aiming for the necessity of more work to address this issue.

Acknowledgements

The authors wish to acknowledge the "Consejería de Medio Rural y Pesca de Asturias, España" for permissions to obtain biological samples. Thanks also to Anadon N. from the Departamento de Organismos y Sistemas, Universidad de Oviedo, who carried out morphological classifications.

References

[1] S. A. Ridgway, D. G. Reid, J. D. Taylor, G. M. Branch and A. N. Hodgson, "A cladistic phylogeny of the family Patellidae (Mollusca: Gastropoda)". *Proc. R. Soc.Lond.,* B., vol. 353, pp. 1645-1671, 1998.

[2] V. Koufopanou, D. G. Reid, S. A. Ridway and R. H. Thomas, "A molecular phylogeny of the Patellidae limpets (Gastropoda: Patellidae) and its implications for the origins of their antitropical distribution". *Mol. Phyl. Evol.,* vol. 11, pp. 138-156, 1999.

[3] A. Sá-Pinto, M. Branco, J. D. Harris and P. Alexandrino, "Phylogeny and phylogeography of the genus *Patella* based on mitochondrial DNA sequence data". *J. Exp. Mar. Biol. Ecol.,* vol.

325, pp. 95-100, 2005.
[4] A. Mauro, M. Arculeo and N. Parrinello, "Morphological and molecular tools in identifying the mediterranean limpets *Patella caerulea*, *Patella aspera* and *Patella rustica". J. Exp. Mar. Biol. Ecol.,* vol. 295, pp.131-143, 2003.
[5] J. P. Thorpe, A. M. Solé-Cava and P. C. Watts, "Exploited marine invertebrates: genetics and fisheries". *Hydrobiologia*, vol. 420, pp. 165-184, 2000.
[6] P. D. N. Hebert, A. Cywinska, S.L. Ball, J. R. DeWaard, "Biological identifications through DNA barcodes". *Proc. R. Soc. Lond.* B Biol. Sci., vol. 270, pp. 313-321, 2003.
[7] O. Folmer, M. Black, W. Hoeh, R. Lutz and R. Vrijenhoek, "DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates". *Molecular Mar. Biol.Biotech.*, vol. 3, pp. 294-299, 1994.
[8] J. P. Wares and C. W. Cunningham, "Phylogeography and historical ecology of the North Atlantic intertidal". *Evolution,* vol. 12, pp. 2455–2469, 2001.
[9] S. J. Hawkins, H. B. S. M. Côrte-Real, F. G. Pannacciulli, L. C. Weber and J. D. D. Bishop, "Thoughts on the ecology and evolution of the intertidal biota of the Azores and other Atlantic islands". *Hydrobiologia*, vol. 440, pp. 3-17, 2000.
[10] F. Espinosa and T. Ozawa, "Population genetics of the endangered limpet *Patella ferruginea* (Gastropoda: Patellidae): taxonomic, conservation and evolutionary considerations". *J. Zool. Syst. Evol. Res.*, vol. 44, pp. 8-16, 2006.
[11] E. Fischer-Piette and J. Gaillard, "Les Patelles au long des cotes Atlantiques Ibériques et nord Marocaines". *J. Conchyliologie*, vol. 99, pp. 135-200, 1959.
[12] M. P. Miyares, Biologia de *Patella intermedia* y *P vulgata* (Mollusca, Gasteropoda) en el litoral asturiano (N. de España) durante un ciclo anual (Diciembre de 1978 a Noviembre 1979). *Bol. Cienc. Nat.* I.D.E.A., vol. 26, pp. 173-192, 1980.
[13] J. A. Ortea, "El género *Patella* Linné 1758 en Asturias". *Bol. Cienc. Nat. I.D.E.A.,* vol. 26, pp. 57-72, 1980.
[14] F. P. Lima, N. Queiroz, P. A. Ribeiro, Hawkins and A. M. Santos, "Recent changes in the distribution of a marine gastropod, *Patella rustica* Linnaeus, 1758, and their relationship to unusual climatic events". *J. Biogeogr.,* vol. 33 (5), pp. 812-822, 2006.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 289-294.
ISBN 978-88-8303-295-0. EUT, 2010.

Molecular Identification of Italian Mouse-eared Bats (genus *Myotis*)

Andrea Galimberti, Adriano Martinoli, Danilo Russo,
Mauro Mucedda, Maurizio Casiraghi

Abstract — Despite the fact that the genus *Myotis* (Mouse-Eared bats) is one of the most investigated microchiropteran groups, recent molecular studies highlighted the presence of several cryptic species with substantial implications for ecological and conservation issues. Our dataset includes 55 *coxI* sequences from 11 morphologically-identified Italian Mouse-Eared bats species. We applied an integrated approach comparing data from a traditional morphological identification and molecular variability in a fragment of the mitochondrial *coxI* gene (DNA barcoding). Our results clearly show a strong coherence between the two identification approaches for almost all of the examined species, and revealed interesting patterns of intraspecific variability within the species *M. nattereri*. Finally, we successfully tested the efficacy of our identification method on undetermined individuals sampled in the field.

Index Terms — cryptic species, DNA barcoding, integrated taxonomy, Vespertilionidae.

1 Introduction

Mammals are usually considered as one of the best-known animal groups. However, several studies provided clear evidences that bats (order Chiroptera) are characterized by a high incidence of overlooked taxa due to their cryptic morphology and habits [1]. A clear example of this situation is given by the recent taxonomic changes within the family Vespertilionidae (one of the best-studied taxonomic group of bats) in the Western Palearctic. Thanks to the development of recent molecular techniques, the number of species within

A.Galimberti and M. Casiraghi are with the Department of Biotechnologies and Biosciences, University of Milan-Bicocca, Milan, Italy. E-mail: tgalimba@gmail.com.
A. Martinoli is with the Department of Environmental Health and Safety, Universities of Insubria-Varese, Italy. E-mail: adriano.martinoli@uninsubria.it.
D. Russo is with the Department Ar bo Pa Ve, Laboratory of Applied Ecology, Agrarian faculty, University of Naples "Federico II", Naples, Italy. E-mail: danrusso@unina.it.
M. Mucedda is with the Centro per lo Studio e la Protezione dei Pipistrelli in Sardegna,Sassari, Italy.

this family has risen from 37 to 54, with at least 8 new cryptic species identified in Europe [2], [3], [4].

Within the Mediterranean basin, the Italian peninsula is one of the most important biodiversity hotspots for bats and other taxa [5]. It has been hypothesized that this peninsula would have provided stable habitats during ice ages, where species survived leading to the generation of new cryptic lineages [5]. 33 microchiropteran species out of the almost 40 currently known to live in Europe are reported for Italy [5], the family Vespertilionidae being the most diverse and abundant (8 genus and 27 species). As showed in recent studies, the family Vespertilionidae is characterized by high levels of cryptic diversity [3], [4] and in particular in Italy, at least 11 different species are included in the group of Mouse-eared bats (genus *Myotis*). This taxon is the most problematic concerning species identification, due to the presence of cryptic species like in the 'Whiskered-bats' complex (i.e.: *M. mystacinus*, *M. Brandtii*, *M. alcathoe*), peculiar biogeographical histories (i.e.: *M. Myotis* and *M. Blythii*; [6]) and genetically-uncharacterized lineages (i.e.: *M. Nattereri*; [2], [4]).

Despite the fact that the genus *Myotis* includes several threatened species, the compilation of realistic action plans for their conservation is biased by some practical difficulties: bats are hard to observe because of their elusive habits. Moreover, several species are cryptic and sometimes it is impossible to reach a correct identification in the field, especially for juveniles or females [5], [7]. Despite the fact that morphological identification keys are available for European bats (e.g. [7]), integration with molecular approaches has proven to be efficient in detecting morphologically cryptic species [2], [3], [4], [8], [9]. An efficient and widely used molecular tool in species identification is DNA barcoding [10]. This technique is based on the analysis of the variability in the nucleotide sequence of a short, standardized region of the genome (among metazoans the 5'- end of the mitochondrial subunit 1 of cytochrome c oxidase), to evaluate differences among species [10]. A few studies have shown the efficacy of *coxI* in identifying bats species, but any (e.g.:[1]) work was conducted with a standardized DNA barcoding approach for European and Italian *Myotis* and other Vespertlionids.

The main objectives of our study are: i) to compile a reference dataset of *coxI* sequences from all the Italian *Myotis* species; ii) to test the coherence between a molecular approach and the morphologically-based taxonomy and iii) to investigate the intraspecific molecular variability of the barcode region to reveal the presence of undescribed cryptic lineages.

2 Materials and Methods

2.1 Sampling, DNA extraction, sequencing and alignment

The samples analysed in this study derive from 55 bats belonging to 11 *Myotis* species collected during 2006-2007 from 16 Italian localities distributed along the Italian peninsula. Bats were identified by researchers of the GIRC (Italian Chiroptera Research Group). Tissue samples (i.e.: 'punches' of patagium, 3mm large) were stored in ethanol 96%. According to the protocol specified by the

Biorepositories initiative (http://www.biorepositories.org) each sample was vouchered with the id Institution name 'MIB:ZPL:' followed by a progressive numeric code. Sixteen unidentified individuals belonging to the genus *Myotis* were also collected.

Total genomic DNA was extracted using a guanidinium thiocyanate and diatomaceous earth protocol [11]. *coxI* amplification and sequencing were obtained following the laboratory protocols provided by [1]. Sequences were checked and aligned following the approach described in [12] and, after checked for the presence of pseudogenes and numts (i.e. nuclear mitochondrial pseudogenes), alignment was cut to 560 bp in order to have all the sequences of the same length.

2.2 Molecular dataset definition and DNA barcoding analyses

To evaluate the efficacy of a DNA barcoding approach as a molecular tool to identify Italian Mouse-Eared bats species, we assembled a 'Reference Dataset' including only *coxI* sequences from Italian morphologically identified bats belonging to the genus *Myotis*. An 'Optimum Threshold' value of molecular divergence (OT) was then calculated (following [12]) directly from the whole range of molecular variability of the molecular dataset. OT value maximizes the coherence between the morphological identification and the molecular divergence minimizing, at the same time, the total amount of possible mismatches (see Minimum Cumulative Error analysis in [12]). To avoid biases during OT calculation, we removed *Myotis nattereri* barcode sequences because of its high intraspecific variability, probably due to the presence of undescribed cryptic lineages (as clearly shown in [2] and [4])

OT was then used to perform a DNA barcoding analysis on the 'comprehensive molecular dataset', containing all reference barcodes and also the *coxI* sequences of the 16 unidentified Italian *Myotis* sp. individuals and the 9 *Myotis nattereri* samples. Intraspecific molecular variability was then analyzed to test the congruence with previously described species. OT has also been used to predict potentially new taxa within the dataset. Finally, a neighbor-joining (NJ) tree (Fig. 1) was generated from the comprehensive molecular dataset using MEGA 4.1 according to the parameters provided in [12].

3 Results and Discussion

Any *coxI* sequence exhibited indels, stop-codons or numts interferences. The 'Reference Dataset' included 46 barcode sequences 560 bp long belonging to all the Mouse-eared bat species distributed in Italy, except for *M. nattereri* (average number of coxI sequences for every single species: 5, standard deviation: 3.05, range: 1-9).

Bioinformatic analyses on the 'Reference dataset' allowed us to infer the following parameters: mean K2P distance within species 0.6% (standard deviation: 0.7%; range: 0% – 2%), mean K2P distance between species 15.8% (standard deviation: 2.8%; range: 2% – 20.1%) and *coxI* overall mean diversity 14% (standard deviation: 1.1%). Concerning the OT calculation, we obtained

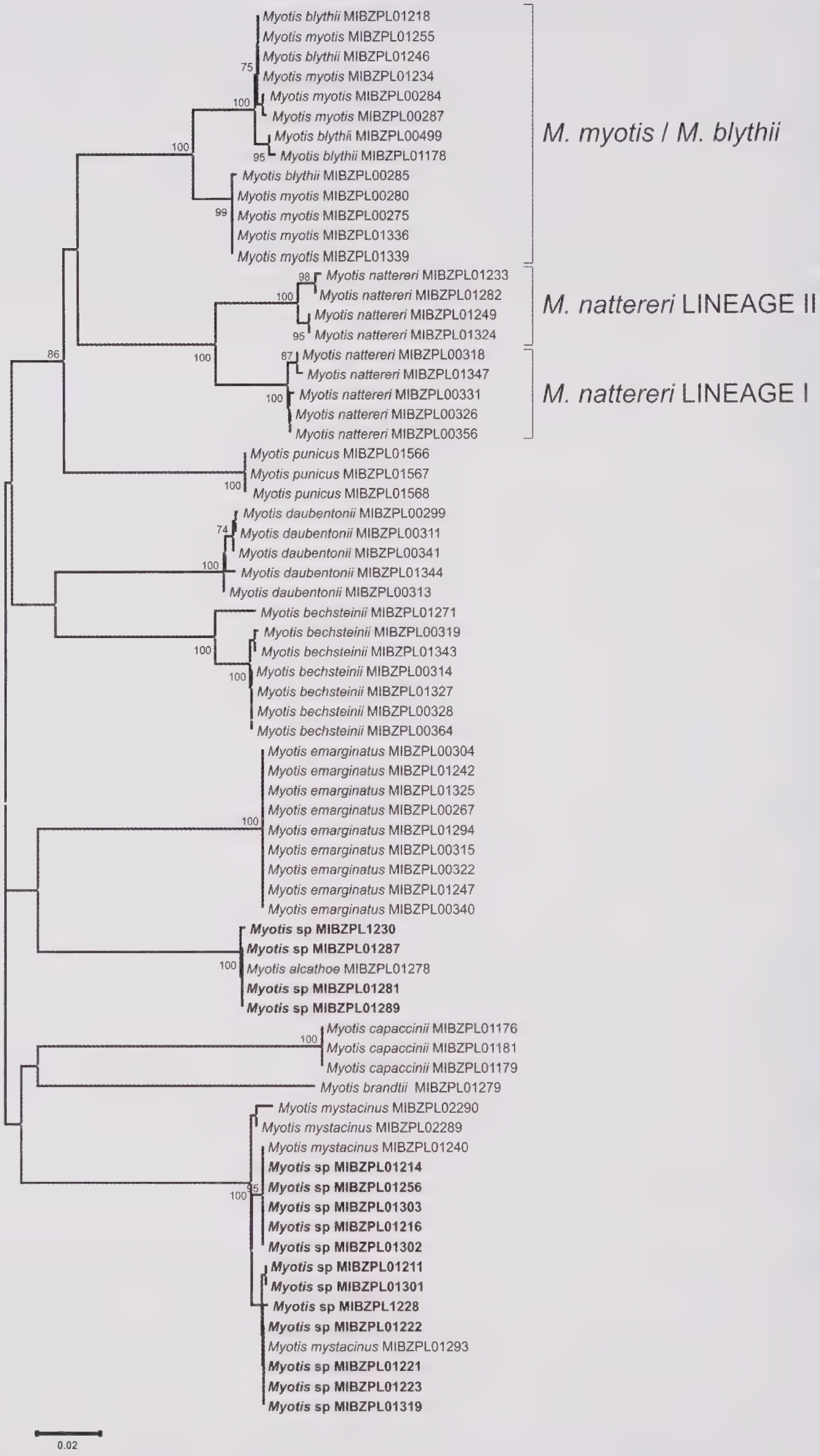


Fig. 1 – NJ Tree. Neighbour joining tree based on *coxI* sequences of Italian Mouse-eared bats generated with MEGA 4.0 (Tamura et al, 2007). Unidentified samples are indicated as "*Myotis* sp.". Cryptic molecular lineages inferred using OT Threshold are indicated with square brackets.

a minimum cumulative error of 4.36% at a Optimum Threshold value of 4.2%. Only one identification mismatch occurred, due to the low interspecific variability (lower than OT) observed between the species *M. myotis* and *M. blythii* (mean K2P distance between species: 2.0%, standard deviation: 1.7%, range: 0% – 4.1%). A similar result has been previously reported in other molecular studies on these taxa (e.g.: [6], [8]) and relies on the fact that a series of introgression events having occurred repeatedly during the recent colonization of Europe by *M. blythii* from Asia. Hybridization is still ongoing in the areas of sympatry (e.g.: in Italy), therefore suggesting an unclear taxonomic status of these taxa in the Western Palearctic.

The application of the OT value on the 'comprehensive dataset' allowed to assign the unidentified specimens (Fig.1) to 12 *M. mystacinus* and 4 *M. alcathoe*. These are two cryptic sympatric species of Mouse-eared bats one of which (*M.alcathoe*) has been recently described [13] and which status in Italy is almost unknown.

Moreover, OT revealed the presence of two cryptic lineages within the taxon *M. nattereri* (here tentatively named 'Lineage I' and 'Lineage II') exclusive of Northern and Southern localities of the peninsula respectively (Fig.1). The mean K2P distances within each lineage are 0.7% and 0.4% for Lineage I and II respectively, while the mean K2P distance between the two lineages is higher than OT: 5.6%. Garcia-Mudarra and colleagues [4] recently identified at least four European cryptic molecular lineages within this taxa and they concluded that ecological as well as morphological studies would be desirable before any definitive conclusions can be drawn about its taxonomic status. Moreover, preliminary molecular comparisons among our lineages and other mitochondrial sequences available in GenBank (i.e: ND1 and cytb) revealed that the Southern Italian 'Lineage II' discovered by our DNA barcoding approach is completely undescribed and could represent a new cryptic *Myotis* species for the Western Palearctic (data not shown).

4 Conclusions

Our study provides clear evidences that DNA barcoding is a reliable and efficient tool for the discrimination of almost all the Italian Mouse-eared bats, showing a high strength of coherence between data based on classical morphology and variability in the mitochondrial *coxI* barcode region. OT value calculated from our dataset allows to infer a clear taxonomic assignment for all the morphologically-unidentified individuals collected in the field. Moreover, the OT value inferred from the molecular dataset is efficient to reveal the presence of undescribed cryptic lineages within known species, like the case of *M. nattereri*. These results suggest that DNA barcoding could be successfully used as a reliable support to ecological studies in order to develop efficient conservation strategies for endangered bats populations.

Acknowledgement

This work was in part supported by a grant from Fondazione Cariplo.

References

[1] E. L. Clare, B. K. Lim, M D. Engstrom, J. L. Eger and P. D. N. Hebert, "DNA barcoding of Neotropical bats: species identification and discovery within Guyana", *Molecular Ecology Notes*, vol. 7, pp. 184-190, Mar 2007.

[2] C. Ibanez, J. L. Garcia-Mudarra, M. Ruedi, B. Stadelmann and J. Juste, "The Iberian contribution to cryptic diversity in European bats", *Acta Chiropterologica*, vol.8, pp. 277-297, 2006.

[3] F. Mayer, C. Dietz and A. Kiefer, "Molecular species identification boosts bat diversity", *Frontiers in Zoology*, vol. 4(1), p. 4, 2007.

[4] J. L. Garcia-Mudarra, C. Ibanez,and J. Juste, "The Straits of Gibraltar: barrier or bridge to Ibero-Moroccan bat diversity?", *Biological Journal of The Linnean Society*, vol. 96, pp. 434-450, Feb 2009.

[5] P. Agnelli, A. Martinoli, E. Patriarca, D. Russo, D. Scaravelli and P. Genovesi (eds.), "Guidelines for bat monitoring: methods for the study and conservation of bats in Italy", Min. Ambiente – Ist. Naz. Fauna Selvatica, Rome and Ozzano dell'Emilia (Bologna), Italy, *Quad. Cons. Natura* Series, vol. 19bis, 2006.

[6] P. Berthier, L. Excoffier and M. Ruedi, "Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*", *Proc. R. Soc. B.,* vol. 273, pp. 3101-3109, 2006.

[7] C. Dietz and O. von Helversen, "Illustrated identification key to the bats of Europe", available at http://www.fledermaus-dietz.de/publications/publications.html, version 1.0., Dec. 2004.

[8] F. Mayer and O. von Helversen, "Cryptic diversity in European bats", *Proc. R. Soc. B.,* vol. 268, pp. 1825-1832, Sept. 2001.

[9] S. M. Goodman, C. P. Maminirina, N. Weyeneth, H. M. Bradman, L. Christidis, M. Ruedi and B. Appleton, "The use of molecular and morphological characters to resolve the taxonomic identity of cryptic species: the case of *Miniopterus manavi* (Chiroptera, Miniopteridae)", *Zoologica Scripta*, vol. 38(4), pp. 339-363, Jul. 2009.

[10] P. D. N. Hebert, S. Ratnasingham and J. R. de Waard, "Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species", *Proc. R. Soc. B.,* vol. 270, suppl. 1, pp.S96-S99, Aug. 2003.

[11] U. Gerloff, C. Schlötterer, K. Rassmann, I. Rambold, G. Hohmann, B. Fruth and D. Tautz, "Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living bonobos (*Pan paniscus*)", *Molecular Ecology*, vol. 4, pp. 515-518, 1995.

[12] E. Ferri, M. Barbuto, O. Bain, A. Galimberti, S. Uni, R. Guerrero, H. Ferté, C. Bandi, C. Martin and M. Casiraghi, "Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda)", *Frontiers in Zoology*, vol.6(1), Jan 2009.

[13] O. von Helversen, K. G. Heller, F. Mayer, A. Nemeth, M. Volleth and P. Gombkötö, "Cryptic mammalian species: a new species of whiskered bat (*Myotis alcathoe* n. sp.) in Europe", *Naturwissenschaften*, vol.88, pp. 217-223, May 2001.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 295-299.
ISBN 978-88-8303-295-0. EUT, 2010.

Identifying algal symbionts in lichen symbioses

Martin Grube, Lucia Muggia

Abstract — Lichens are a ubiquitous terrestrial symbiosis of fungi with photoautotrophic microorganisms. The identification of the hosted photoautotrophs is notoriously difficult. Molecular data to clarify evolutionary relationships on the involved algal and cyanobacterial lineages are accumulating, but the assignment to species is challenging for various reasons. One of the challenges is the limited knowledge on the alpha diversity of photoautotrophs. New lineages are being discovered with increasing amounts of sequencing. Identification tools could incorporate these aspects, by routinely updating the assignment process. We propose the establishment of a classification tool using algal sequence data from public databases.

Index Terms — lichens, symbionts, photobionts, ITS, actin.

1 Introduction

Lichens are symbioses of fungi and photoautotrophic partners (algae and/or cyanobacteria). Lichens are widespread in all climatic zones and cover more than 8% of the land surface [1]. Lichens are generally named after the morphology-determining fungal partner which represents more than 18.800 known species of Ascomycetes [2]. Contrarily, the knowledge about photobiont species diversity is still limited. The determination of lichen photobionts is complicated due to the lack of diagnostic characters for routine analyses. Algae in lichenized stage do not express useful characters at all, and cultivation of algae is time-consuming and not yet possible for some lineages [3].

Recent DNA sequence analyses have studied phylogenetic diversity of algal partners in lichen symbioses. About 50% of the lichen fungal species associate with single-celled green algae, and most of these belong to the genus *Trebouxia* (Trebouxiophyceae, Chlorophyta). Although morphologically similar, different genetic lineages of these photobionts are detected in wide geographic ranges of the same lichen fungal species.

Algal symbiont sequence information is usually obtained by using algal specific primers for amplification from total lichen (=holobiont) extracts, which avoids the contamination by sequences of the fungal partner, and multiple co-occurring bacteria. The phylogenetic analyses of the internal transcribed spacer

The authors are with the Institut für Pflanzenwissenschaften Karl-Franzens-Universität Graz Holteigasse 6, 8010 - Graz, Austria. E-mail: martin.grube@uni-graz.at.

(ITS) nuclear region uncovered relationships among trebouxioid photobionts and selectivity of the fungal partners for their algae. Sequence data of this group of algae have accumulated significantly in the past years and meanwhile a search in GenBank using "*Trebouxia* ribosomal ITS" returns 1356 hits (as of 07.07.2010).

Phylogenetic analyses are used to assign sequenced strains to named species. This is not at all a trivial procedure, because sequence divergence among recognized species is far from equal: some taxa are separated by few nucleotide changes while more pronounced sequence divergence can be detected among other species. Thus there is uncertainty about the assignment of sequences within the range of divergence among two species.

Previous study found that sequence divergence of *Trebouxia* is recognized in several main phylogenetic clades, which have been designated by letters A, I, G, S [4]. Within these clades, subclades are distinguished by numerals. This has lead to a fairly resolved phylogenetic classification of lineages in *Trebouxia*. This phylogeny does not agree perfectly with phenotypical classification and species taxonomy. It has also been revealed that diversity of these algae was previously underestimated and includes many yet to be described species. This includes entirely new lineages as well as the better characterization of yet cryptic lineages within broadly understood species. Nevertheless, new names for species are rarely introduced, e.g. when morphological characters correlate with distinct phylogenetic positions. Fig. 1 (modified from [5]) displays the challenges. In a recent publication [3] we could show that the sister clade (*Trebouxia* sp. 1) of *Trebouxia arboricola* cannot be distinguished by ultrastructural data of cultured algae from that species, whereas the distinct clade *Trebouxia* sp. 2 could not be cultured with standard methods. Thus the phylogenetic suggest a distinct clade but phenotypic support is still missing. Whether the basal lineages of that clade could represent a further species still needs to be awaited and supported by further sequence data.

Variation within an algal species will more precisely be estimated with more sequence data. Because sexual stages are cryptic in lichenizing *Trebouxia*, sequence evolution in clonal lineages could blur species delimitation. We expect that species are increasingly recognizable as 'clusters' in the sequence space with appropriate gene loci. Sequence divergence of ITS is suitable for DNA barcoding of green algal lichen symbionts. We therefore suggest establishing an automated assignment tool that tests query sequences against a regularly updated database of lichen algal ITS sequences. Automated classifiers have been incorporated in the RDP database of bacterial rDNA [6], but do not yet exist for eukaryotes. Moreover, assignments have to consider the growing amount of environmental sequences without assignment to taxonomic names. We are therefore exploring methods to assess the coherence of related sequence as clusters in the sequence space, and the confidence of their assignment to species names. This work is still in progress and more details will be presented at the Bioidentify meeting in Paris.

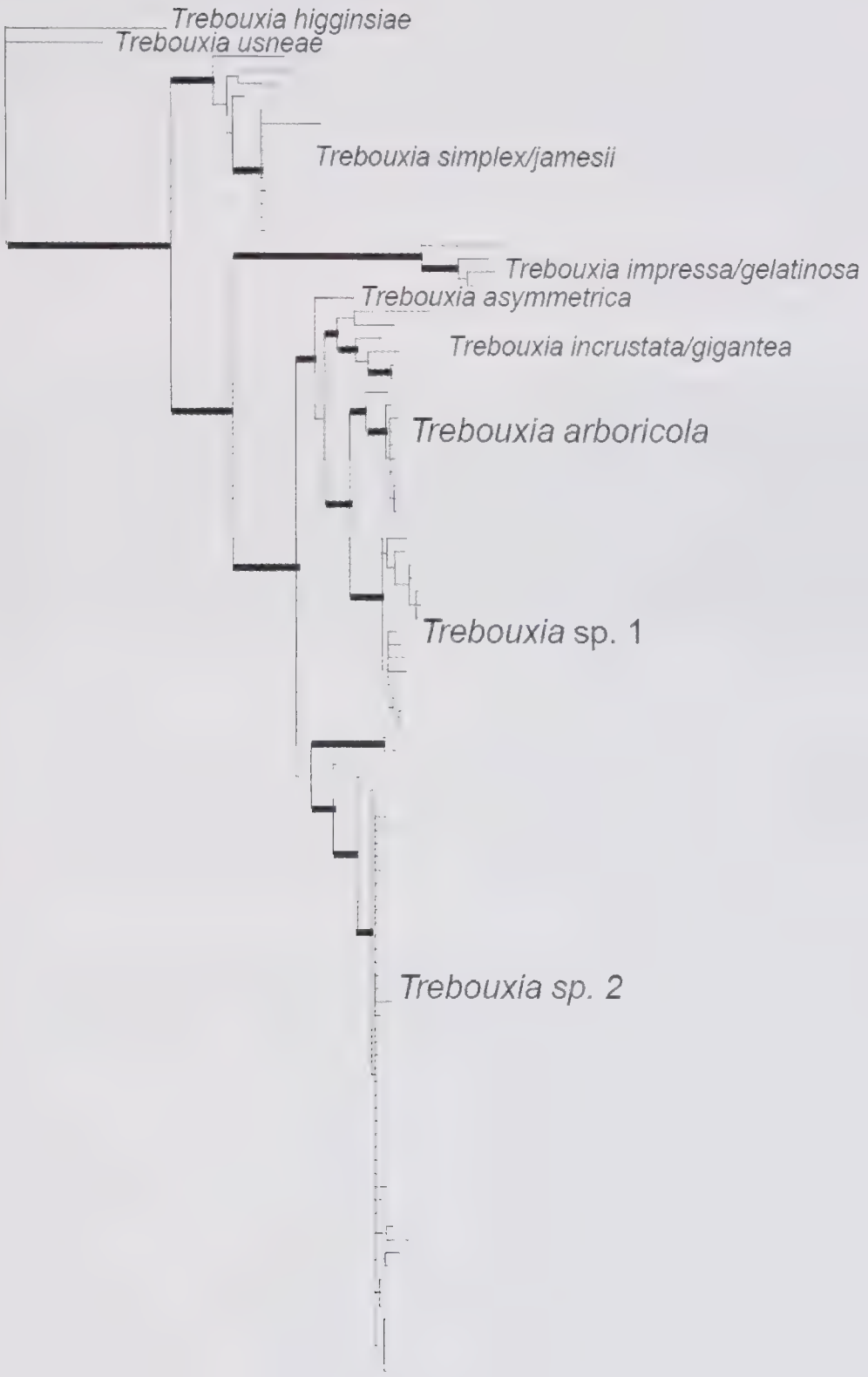


Fig. 1 – Phylogenetic tree of *Trebouxia* species as algal symbionts in lichens (modified from [3]). The tree is constructed using ITS rDNA sequence data. Symbionts in Mediterranean samples of the lichen *Tephromela atra* are named informally as *Trebouxia* sp. 1 and sp. 2. Further information is required for taxonomic recognition of these cryptic species.

2 Multiple phototrophs in lichens

Lichen fungi may internalize more than one algal symbiont. This is clearly observed when a green algal lichen thallus contains nitrogen-fixing cyanobacteria in specialized organs. Less clear, however, are cases where several algae of the same green algal lineages are involved. Culture based studies have previously shown that a single individual of lichen can host several lineages of *Trebouxia*. This variation is often poorly detected with conventional PCR approaches using whole thallus DNA extracts, whereby usually only one distinct sequence

is detected. Any additionally occurring algal sequences are obscured by the exponential amplification of the most common sequence during PCR. Because diagnostic characters of algae in the lichenized stage are hardly available it is still unclear how multiple strains of algae are distributed in lichens. Are additional algae merely epibionts, are they evenly distributed in low abundance throughout the thallus, or are they localized in certain parts of a lichen thallus?

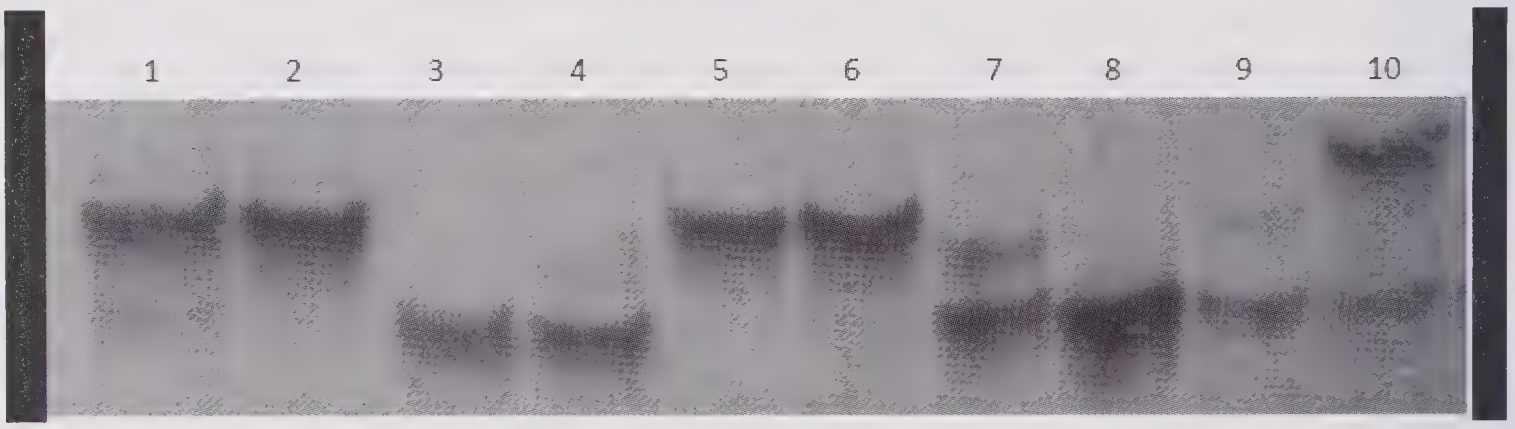


Fig. 2 – Identification of multiple algal symbionts in the lichen *Lecanora muralis* by single strand conformation polymorphism (SSCP) detection. External (odd lanes) and internal (even lanes) areoles from 5 lichen individuals were analysed. Bands of equal position represent distinct algal genotypes. *Lecanora muralis* associates with several algal genotypes, and areoles 1, 7, 9, and 10 display heterogeneity for the algal partner.

Several methods are available to analyse the composition of microbial communities. One of these is single strand conformation polymorphism (SSCP) analysis. With this method single-stranded DNA fragments can be separated on a gel according to their nucleotide sequence variation. The separated bands can then be excised from the gels for sequencing. Each band with different run-length represents a different sequence/strain of algae. Best results are obtained with sequence fragments of 200-300 nucleotides length, i.e. within the size range of the ITS subregions flanking the internal 5.8S rDNA. We are now using this method to explore the algal composition in lichens in greater detail. Fig. 2 shows a detail of a SSCP analysis of individual areoles of 5 specimens of the euryoecious lichen *Lecanora muralis*. Odd lanes are from external areoles, even lanes are from internal areoles. Symbiont heterogeneity is observed among thalli and within thalli (especially in the two samples representing lanes 7-10).

We assume that different strains are not evenly distributed at dissimilar abundances in the entire thallus. We rather think that thallus outgrowths can newly associated with algae, and that different algae could be detected in different areoles or lobes of a lichen individual. We expect that this flexibility could contribute to the ecological adaptivity of euryoecious lichens.

3 Conclusion

The ITS rDNA sequences of lichen photobionts are useful DNA barcodes to study partner selectivity in symbioses. Here we focused on algal partners in lichens. The major challenge of this work is the still unsettled taxonomy of algae, and that several algal symbionts may be present in lichen individuals. A self-organising classification tool that uses regularly updated sequence information

on algal lichen symbionts is under development.

Acknowledgement

We thank Toby Spribille (Graz) for comments on the text. This work was supported in part by a grant from FWF (P17601).

References

[1] J. V. Ahmadjian, "Lichens are more important than you think". *Bioscience* vol. 45, pp. 123–124, 1995.

[2] T. Feuerer and D. L. Hawksworth, "Biodiversity of lichens, including a world-wide analysis of checklist data based on Takhtajan's floristic regions." *Biodiversity and Conservation*, vol. 16, pp. 85–98, 2007.

[3] L. Muggia, G. Zellnig, J. Rabensteiner and M. Grube, "Morphological and phylogenetic study of algal partners associated with the lichen-forming fungus *Tephromela atra* from the Mediterranean region". *Symbiosis*, on line first, 2010.

[4] G. Helms, *Taxonomy and Symbiosis in Associations of Physciaceae and* Trebouxia. Univ. Göttingen (Doctoral thesis), 2003.

[5] L. Muggia, M. Grube and M. Tretiach, "Genetic diversity and photobiont associations in selected taxa of the *Tephromela atra* group (Lecanorales, lichenised Ascomycota)". *Mycological Progress*, vol. 7, pp. 147-160, 2008.

[6] J. R. Cole, Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity and J. M. Tiedje, "The Ribosomal Database Project: improved alignments and new tools for rRNA analysis", *Nucleic Acids Research*, vol, 37 (Database issue), D141-D145; doi: 10.1093/nar/gkn879, 2009.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 301-305.
ISBN 978-88-8303-295-0. EUT, 2010.

Identification of polymorphic species within groups of morphologically conservative taxa: combining morphological and molecular techniques

Kim Larsen, Elsa Froufe

Abstract — Identification of small species with high levels of intra-specific polymorphism within groups of inter-specific morphologically-conservative taxa, presents numerous obstacles for biodiversity and ecological studies. This is particularly true for deep-sea studies that often reveal a great number of species but only few numbers of specimens. It is here proposed to deal with such cases by extrapolating information obtained from highly detailed baseline studies. Such baseline studies should include information about sexual and ontogenetic variation and should include a combination of both morphological and molecular techniques.

Index Terms — baseline studies, polymorphism, sibling species, species identification.

1 Introduction

The identification of species can be problematic enough when dealing with taxa which include a large number of morphologically similar species. The obstacles can increase manifolds with smaller taxa that display few stable characters and show tendencies towards reductions. Adding the complications of substantial sexual and ontogenetic variations, the results are often misleading to the point of being meaningless. This is particularly true for deep-sea studies that often reveal numerous species but only few specimens for each species.

One example of such a problematic group is the Tanaidacea (Crustacea: Peracarida), but there are many other, similar difficult taxa, among the smaller invertebrates. In the Tanaidacea species, differentiation is notoriously difficult,

The authors are with CIIMAR (Center for Interdisciplinary Investigation of the Marine environment), LMCEE (Laboratory of Marine Community Ecology and Evolution), Rua dos Bragas 289. 4050-123, Porto, Portugal. E-mail: tanaids@hotmail.com, efroufe@ciimar.up.pt.

and males/juveniles often share no species-specific characters with females; even family level identification of males can be hazardous [1]. In many families, multiple polymorphic males exist- the consequence of a peculiar reproductive strategy involving protogynous hermaphrodites- and this causes additional problems [2]. As males locate females by roaming the substrate, they are exposed to high predation pressure that - combined with their non-feeding life-style - makes the life span of males short. In situations where depletion of males from the population occurs, some females may molt into males at several different instars, each resulting in a morphologically different male (up to four different male morphs have been recorded in one species) [2]. Ontogenetic variations among adult females are also known to cause problems [1].

At the same time, the tanaidaceans are infamous for creating species complexes containing many, often sympatric, species that display a very conservative inter-specific morphology. This again makes species identification exceedingly difficult [1].

Tanaidacea are particularly common in deep-sea substrates, where they constitute a major proportion, up to 22 % by some estimates, of the total fauna (in terms of biodiversity) [3], [4]. It is clearly undesirable for any scientific study (particularly in biodiversity and ecology) that such a large proportion of the fauna cannot be identified, but the solution to this problem is not apparent due to time constraints and lack of available expertise. Many large scale biodiversity programs rely heavily on cheap (and poorly trained/supervised) student help to process the often enormous material of small benthic invertebrates. Clearly, given the problems inherent with taxa displaying such troublesome attributes as described above, such personnel have little chance of successful identification (we have personally observed as much as 50% misidentifications in collections from biodiversity studies).

2 Methods

The methods we suggest here for species identifications are not new but make use of an expanded procedure. Firstly the samples should be screened and identified to order. Thereafter, samples with large number of specimens should be given priority. Those high-value samples should then be sent to taxonomical experts for 'baseline' processing. Do NOT use untrained student assistants for this part. Once the experts have reported the baseline study, make the identifying personnel use these for comparisons with each single species. Singletons should not be dissected, but not assigned species rank either (like for example '*Sp. A*'), until comparisons have been made with other singletons of different instars that may belong to the same species.

2.1 Morphological Methods

The baseline study should be conducted in the utmost details, including the dissection, description, and illustration of ANY appendages (not just those normally regarded as taxonomically informative) for ALL developmental stages encountered. Both the lateral and dorsal view of the body should be

drawn. Dissection should include appendages from BOTH sides of the body. Appendages should be mounted in glycerin dyed with clorasol black, sealed with nail polish, and stored for further studies. All character transformations seen from manca-juvenile-adults should be noted and illustrated as a transformation diagram. The baseline study should result in the manufacturing of a guide to identification of genders and developmental stage, and supplied to the personal conducting the identifications of the entire material.

2.2 Molecular methods

2.2.1 DNA Extraction

The main problem with DNA extraction from such specimens is the very low yield of starting tissue available (for the smaller taxa, the entire animal has to be used, since a leg or other appendages do not yield enough DNA). Therefore, the extraction is crucial for further analyses and usually requires some modifications to frequently used protocols. There are several DNA isolation techniques. Here we describe our modifications to one frequently used protocol: silica columns. The most crucial points are as follows: VERY thorough grinding of samples, prolonged periods in the several steps stated in the DNA extraction Kit (we use JETQUICK Tissue DNA kit) and also prolonged periods for the final elution step.

Insufficient disruption of starting material leads to low yield and purity, therefore this step is crucial; we use hand-made hard-plastic cylinders which are efficient in disruption and homogenization of the hard crustaceans exoskeleton and also-because of their small size- can be used in micro-centrifuges tubes avoiding the risk of contamination (they can also be autoclaved) and avoiding loosing tissue (the same micro-centrifuges tubes can be used for proteinase-k digestion).

Extraction can be performed according the JETQUICK protocol but should be modified by increasing the time length of each step, from incubation with proteinase K to each centrifuge step (we used double time). Due to the low final DNA concentration; the same elution solution should then be used for the DNA elution and the same for the second elution step (pre-warmed at 70°C for five minutes). Densitometric measurements are not useful for detection of small amounts of DNA [5] so the “Qubit” flurometer is ideal (requires only 1µl DNA elution).

2.2.2 PCR

The basic “PCR rules” HAVE to be employed when dealing with these kind of samples, e.g., cleaning the bench top with alcohol before setting up reactions, using plugged tips for all PCR reagents (to avoid contamination), always including a sample without template as a negative control to check for contamination of the reagents. The most crucial points are as follows: short length of PCR products (optimum of 300-350 bp) and higher number of PCR cycles.

The amount of DNA used will depend on the concentration of the sample. It is best to use a “hot start” *Taq* that will provide increased sensitivity, specificity

and yield. Due to the high numbers of PCR cycles needed the quality of the *Taq* is also important (we used Platinum *Taq* DNA Polymerase). Finally in order to avoid adding enzyme inhibitors that may be present, we recommend the use of a high PCR final volume (20 µl).

2.2.3 DNA Sequencing

At this stage, the products must be checked for both quantity and quality. Agarose gel electrophoresis can be used to visualize the amount and size of DNA fragments present in the sample, and since usually the amount of final PCR concentration is low when using these type of samples, we recommend to dry up the total PCR product (use a vacuum centrifuge) into a loading agarose gel volume and excised the PCR gel band. We used several different commercial Gel extraction kits, with no significant results among them. The only modification is the final elution step, which should be no higher than 10µl (we used 5 µl). The DNA sequencing can proceed as usual hereafter.

3 Discussion

Given the large material of 'difficult taxa' often encountered during biodiversity/ ecological studies (particularly from deep-sea environments), the limited expertise available on many such taxa, and the financial restraints, it is not possible to have specialists processing all the material. Therefore we propose to deal with these problems by extrapolating information obtained from the highly detailed baseline-studies described above. We are not so much suggesting new 'methods' for species identifications, but rather a different overall procedure of dealing with large amounts of small troublesome taxa. Instead of dealing with samples from one end to the other, we suggest discriminating between samples of 'low' and 'high value', the latter to be dealt with in great details by specialist, and with priority over 'low' value samples. High value samples are those which contain lots of specimens. Particularly deep-sea collections often reveal many species but few specimens and thus offer only few such targets for detailed studies of inter-specific variation. However, due to the patchy distribution often encountered in the deep sea, a few samples (maybe 1 in 100) will contain lots of specimens and most often these will belong to one or two species. These are the samples worth their weight in gold, and those species of which much material exist should be examined (and the species described/redescribed) in great detail, including dissections, illustrations and descriptions of several individuals, of several developmental stages, and of both sexes. At the same time specimens (males, females, juveniles, and mancae) should be processed for molecular studies to verify con-specificity with absolute certainty. Since most families have not been studied in such detail, these baseline studies are needed to provide the detailed information required for processing other species of the same phylogenetic groups but encountered in fewer numbers during the specific survey. Once such a baseline study has been made, other members of the same family can be processed 'normally' by comparing the characters of whatever instar or gender with the information provided by the baseline study.

If an adult female singleton is encounter, then it can be compared with an adult female from the baseline study; if a manca is encounter, it can be compared with a mancae from the baseline study and so forth. We will thus have the information at hand which is needed for firstly correctly identifying the actual instar/gender, and, secondly, for species identification knowing now which characters are stable or not. These may well vary between higher taxonomical groups but are likely to be similar (or more similar at least) within phylogenetically close groups.

We would like to end this paper with a note on descriptions of new taxa. The senior author recently participated in a workshop regarding the description of peracaridean crustacean. The participants received the following request by the one of the organizers:

"We recently collected several thousand deep-sea species of which we estimate half of them to be new to science. We would like to describe these new species but it is a monumental task that we just don't have the time for. We would therefore like the participants of this workshop to come up with some guidelines to how to describe 'bulk' new species in a short abbreviated and timely fashion".

After a short debate the participants unanimously came up with the only possible answer:

"Please don't do that!".

While the person in charge of this overwhelming material only had good intentions and was indeed faced with an impossible task, abbreviated descriptions can only lead to chaos. If descriptions of such small and difficult creatures are to have any value what so ever- now and in the future- it is absolutely paramount that new species are described thoroughly and in minute details.

References

[1] K. Larsen, "Morphological and Molecular Investigation of Polymorphism and Cryptic Species in Tanaid Crustaceans: Implications for Tanaid Systematics and Biodiversity Estimates". *Zoological Journal of the Linnean Society,* vol. 131(3), pp. 353–379, 2001.

[2] J. Sieg, "Evolution of Tanaidacea". In: F. R. Schram (ed.) *Crustacean Issues* 1, *Crustacean Phylogeny,* Rotterdam, 1983.

[3] T. Wolff, "Diversity and composition of deep-sea benthos". *Nature*, London, vol. 267, pp. 780–785, 1977.

[4] K. Larsen, *Deep-Sea Tanaidacea (Peracarida) from the Gulf of Mexico*. Crustaceana Monographs. (Brill, Leiden), vol. 5, p. 387, 2005.

[5] U. M. Csaikl, M. Bastian, R. Brettscheider, S. Gauch, A. Meir, M. Schauerte, F. Scholz, C. Sperisen, B. Vornam and B. Ziegenhagen, "Comparative analysis of different DNA extraction protocols: A fast universal maxi preparation of high quality plant DNA for genetic evaluation and phylogenetic studies". *Plant Molecular Biology Reporter*, vol. 16(1), pp. 69–86, 1998.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 307-313.
ISBN 978-88-8303-295-0. EUT, 2010.

Coffee species and varietal identification

Patrizia Tornincasa, Michela Furlan, Alberto Pallavicini, Giorgio Graziosi

Abstract — There are serious economical reasons to pretend warranties in coffee species and varieties authenticity. Arabica adulteration with Robusta coffees, intentional or not, is carried out at different steps of the coffee chain, from plantation to beverage. We present a method based on a real-time PCR technique to perform: a) a *qualitative* analysis to evaluate the presence/absence of a species in a sample; b) a *quantitative* analysis to amplify Robusta samples only, making possible the detection of less than 5% of Robusta in a mixture. Arabica cultivars sold as specialty coffees can be identified and certified by fingerprinting using SSR markers.

Index Terms — adulteration, *Coffea arabica*, *Coffea canephora*, microsatellites, multiplex PCR, real time PCR, roasted coffee, traceability.

1 Introduction

Coffee is one of the most important products in the international market. The annual consumption exceeds 5 billions kilograms, wich corresponds to 500 billions cups. The genus *Coffea* contains more than 100 species, only two of which, *Coffea arabica* (known as Arabica coffee) and *C. canephora* (known as Robusta coffee) are commercially cultivated. Arabica produces high quality coffee compared to Robusta, and contributes with about 70% of the total world coffee production, being consequently sold at 2-3 times higher prices. Also, among Arabica coffee some cultivars are considered as specialty coffees with peculiar organoleptic characteristics and a very high commercial value. Thus, there are serious economical reasons to pretend warranties in the authenticity of coffee species and varieties. Arabica adulteration with Robusta coffees can be intentional or not and is carried out at different steps of the coffee chain, from plantation (one or both species can be cultivated by the same producer) to coffee beverage.

Green coffee authentication can be very useful for roasters, while that of roasted coffee (beans or ground) should be very interesting for retailers and consumers.

The authors are with the Department of Life Sciences, University of Trieste, P.le Valmaura, 9, I 34127 Trieste, Italy. E-mail of A.Pallavicini: pallavic@units.it.

The methods to distinguish Arabica from Robusta in coffee blends are presently based essentially on the chemical analysis of compounds such as sterols [1], chlorogenic acid and caffeine [2], fatty acids [3], tocopherol [4], etc, but these do not always give reliable results. Nowadays there is no totally reliable method to guarantee coffee authenticity; the only reference is the production chain of coffee, trusting on what the sellers declare through labels of already packed products, or tasting the drink.

Concerning Arabica varieties, it is not possible to distinguish them from the morphology of the seed or from plant phenotype and agronomy. The only method to assess product quality wich is available for dealers is to roast the seeds and taste the coffee beverage. Importers also cannot know whether the small testing sample effectively corresponds to the many hundreds of coffee bags received afterwards.

In recent years, food forensics requires DNA-based methods for molecular analysis. The aim of this approach is to guarantee authenticity of commercially important foods that can be contaminated accidentally or by fraud. Generally, molecular tecniques based on DNA analysis are more effective and reliable than those considering phenotipic characteristics. In particular molecular markers such microsatellites or SSR (simple sequence repeats) are the most suitable for their features: abundance in eucariotic genomes, high level of polymorphism, codominance, locus specificity, PCR detection and high results reproducibility. SSR are widely used in the characterisation of plant species such as rice [6], potato [7], and wheat [8]

Research on coffee in this field is still at the beginning, and only one method based on PCR-RFLP is available [5]. Here a Real-time PCR based method is described for blend coffee analysis.

2 Materials and Methods

2.1 Discerning Arabica and Robusta

Real time PCR can be used to analyse coffee blends for establishing the relative presence of Arabica and Robusta coffee. This method relies on DNA-based probes which are complementary to target sequences in a region internal to PCR primers. Each probe has a fluorescent reporter at one end and a quencher of fluorescence at the opposite end of the probe. The close proximity of the reporter to the quencher prevents detection of its fluorescence; breakdown of the probe by the 5' to 3' exonuclease activity of the Taq polymerase breaks the reporter-quencher proximity and thus allows unquenched emission of fluorescence, which can be detected after excitation with a laser. An increase in the product targeted by the reporter probe at each PCR cycle therefore causes a proportional increase in fluorescence due to the breakdown of the probe and release of the reporter.

2.2 Qualitative analysis of Arabica and Robusta blends

The method is based on the amplification of both Arabica and Robusta samples through the use of Real Time PCR technology. The identification of these two species is given by tree different probes, each one presenting a specific fluorochrome: one is an universal probe, that recognizes both Arabica and Robusta, index of amplification efficiency; a specific probe for *C. arabica*, that binds a sequence present only in Arabica but absent in Robusta; a specific probe for Robusta, that binds a piece of DNA present in *C. canephora* and absent in Arabica. This method allows to carry out qualitative analysis to evaluate the presence/absence of a species in a sample. Additionally, it can also amplify and detect relative quantities of the two species in coffee blends.

2.3 Quantitative analysis of Arabica and Robusta blends

The second method permits only Robusta species amplification also in Arabica/ Robusta blends, through the use of Real Time PCR technology. The detection of the DNA products is given by an universal probe, that recognizes both species. In addition, to avoid undesired amplification of Arabica, a LNA (Locked Nucleic Acid) oligonucleotide clamp was added. This clamp can hybridize in a DNA region present only in Arabica but absent in Robusta, and does not permit primer annealing. This method inhibits the amplification of the most abundant species (usually Arabica) in favour of the less abundant one (Robusta) that can be present in case of fraudulent contamination. Using this method it is possible to give an estimation of the percentage of Robusta present in a mixture.

2.4 Arabica coffee fingerprinting

DNA was extracted from leaves and seeds of 320 different plants of *Coffea arabica* that constitute the Arabica collection of Laboratory of Genetics (University of Trieste, Italy).

Two multiplex PCR reactions were performed on each sample. The M1 reaction involves 9 couples of primers and allows to amplify at the same time 9 microsatellite loci. The second multiplex, PCR M2, contains 7 couples of primers and allows amplification of 7 loci microsatellites. Moreover, the amplification products of M1 and M2 primers are studied to have amplicons with non-overlapping molecular weights. Since this is not always possible, primers are labelled with different fluorophores to distinguish amplification products with similar size. This permits to mix both M1 and M2 products and to analyze them in a single electrophoretic run through the genetic analyzer. The sequences on which primers contained in M1 and M2 were designed are covered by patent [9]. The advantages of this technique are: PCR easy to perform; possibility to analyze 16 microsatellites using only 2 PCR reactions and one run by the genetic analyzer, saving reagents, time and costs.

3 Results and Discussion

3.1 Qualitative and quantitative analysis of Arabica and Robusta blends

Two methods were developed to distinguish Arabica and Robusta species. The qualitative approach permits immediate verification of the presence of Arabica, Robusta or of both in a mixture. Furthermore, we can approximately estimate the quantity of the two species up to 20 % of Robusta. Fig. 1A and Fig. 1B show the specificity of the two fluorescent probes for Robusta and Arabica coffee, respectively. The universal probe (indicated in Fig. 1A and 1B with number 1) gives an amplification efficiency higher than the other two (numbered with 2 and 3).

A

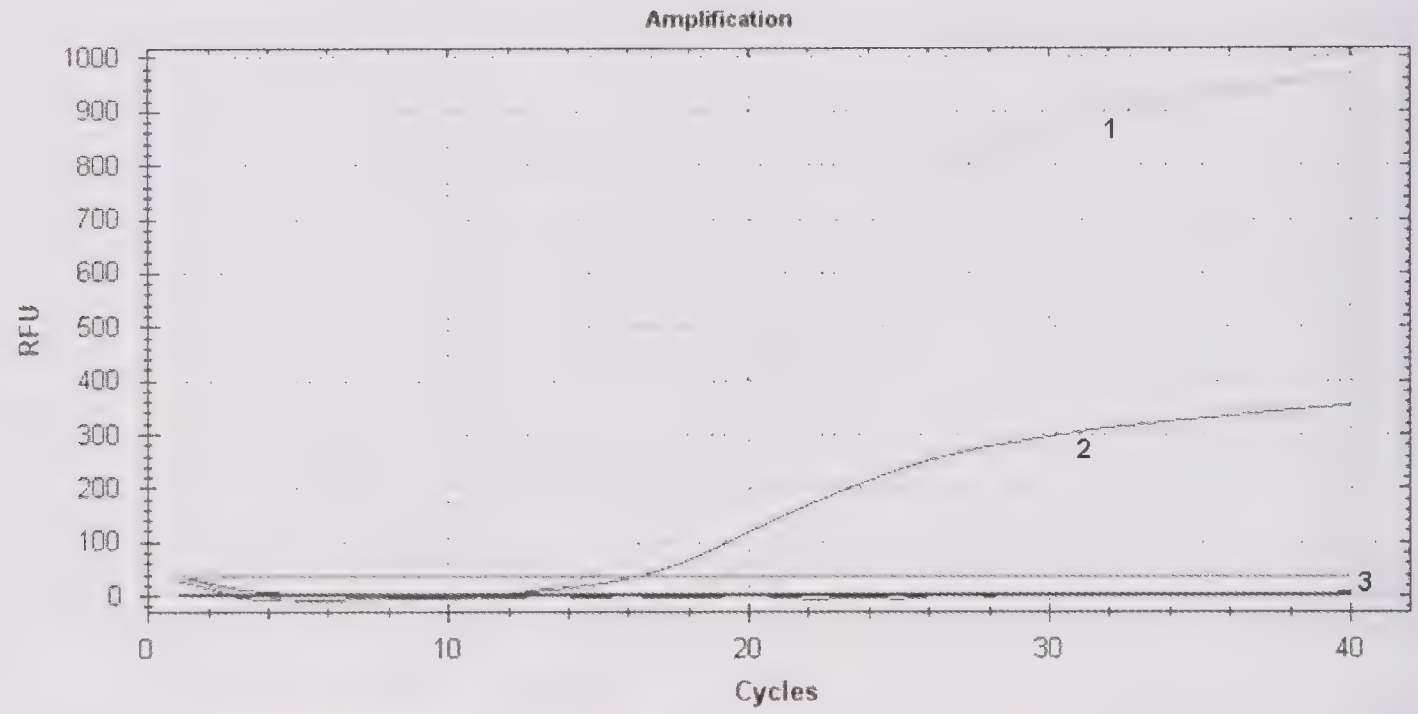


Fig. 1A – Amplification of a Robusta sample. The amplicon detection is given by the universal probe (1) and the *C. canephora* specific one (2), while the Arabica specific probe (3) shows no signal.

B

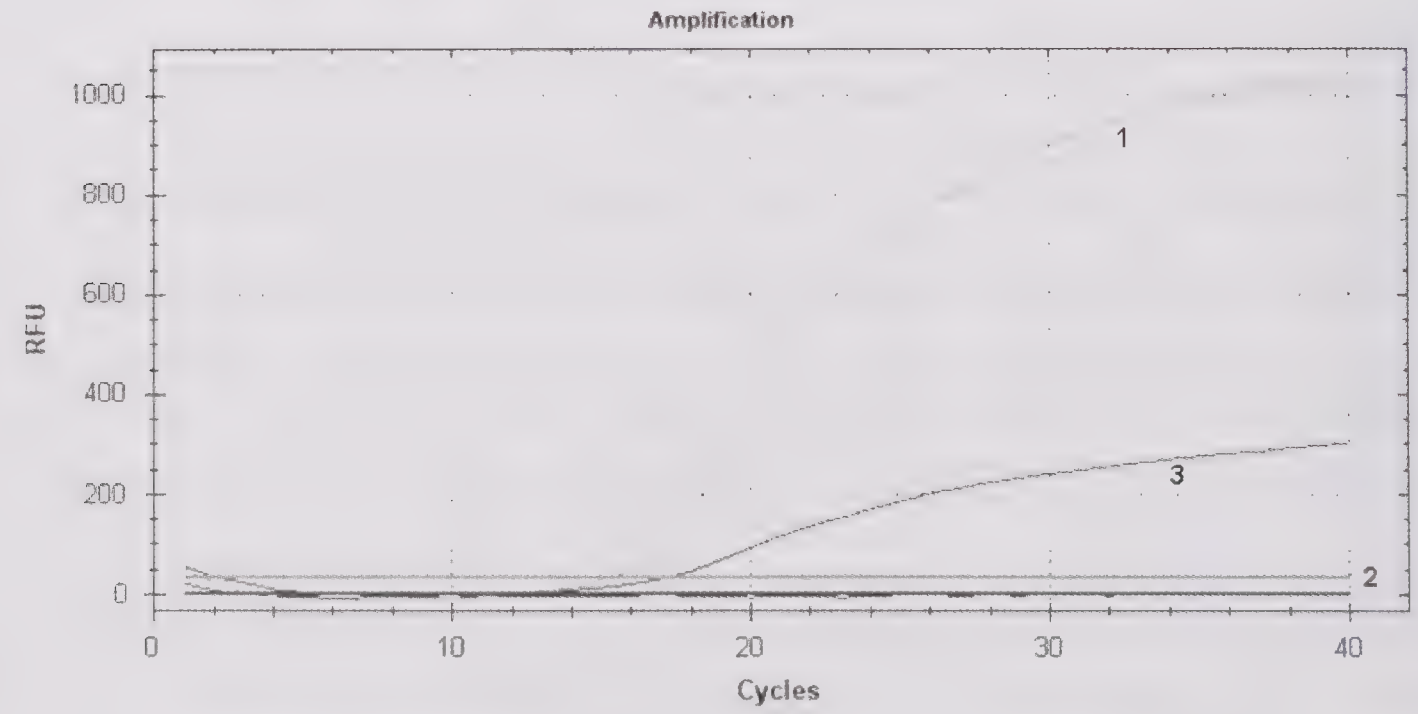


Fig. 1B – Amplification of an Arabica sample; the detection in this case is achieved by the universal probe (1) and the Arabica specific one (3), while the *C. canephora* specific probe (2) gives no signal.

The quantitative method was developed to amplify only Robusta samples, making possible the detection of less than 5% of Robusta in a mixture. The addition of the oligo clamp was required to inhibit Arabica amplification performed by this system. Fig. 2 shows that increasing concentrations of this oligo clamp progressively inhibit amplification of Arabica.

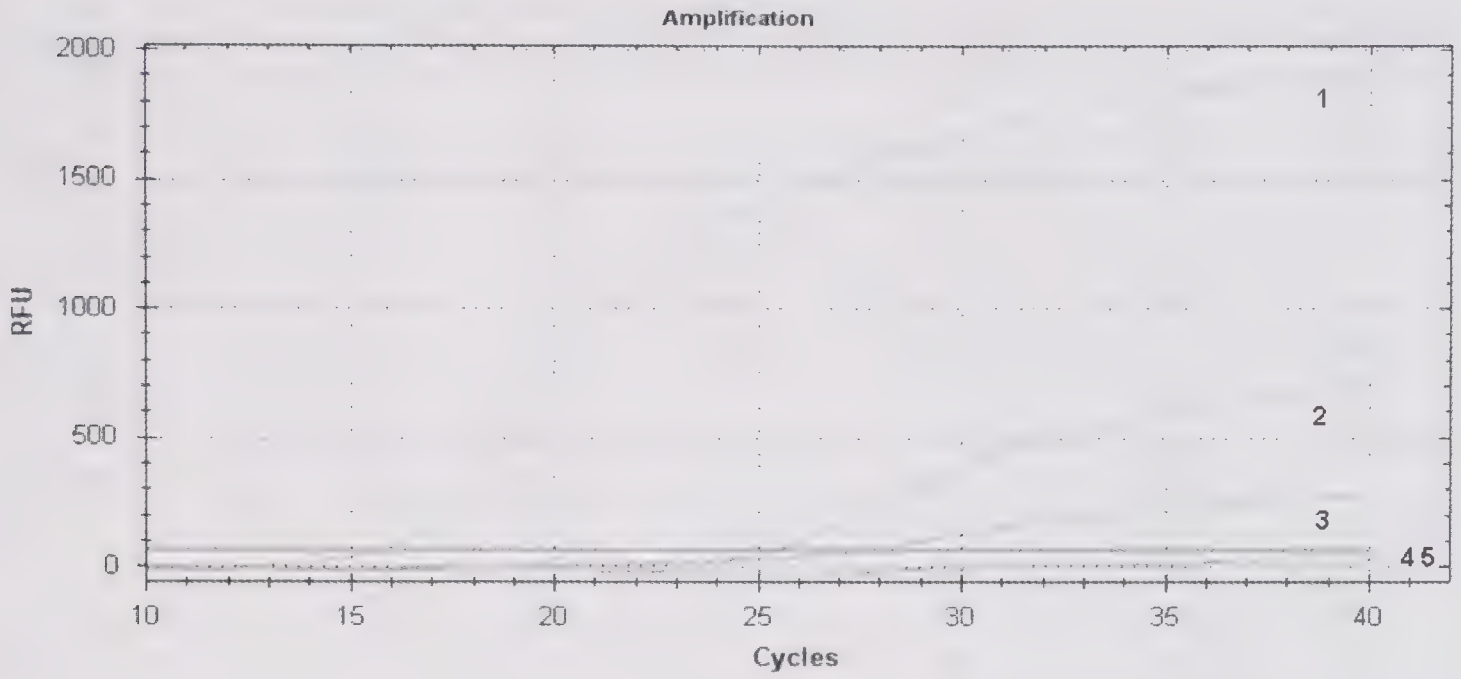


Fig. 2 – Chart showing amplification of Robusta and Arabica with the quantitative system. Curves 1 and 2 represent Robusta and Arabica amplifications, respectively. Curves numbered from 3 to 5 display progressive reduction of Arabica amplification with addition of increasing oligo clamp amounts, in particular: 0,06 µM (3), 0,6 µM (4), 1 µM (5).

Conversely, the oligo clamp doesn't interfere with Robusta amplification (data not shown).

This method can be useful to detect a wide range of Robusta percentages that can be present in a mixture.

3.2 Arabica coffee fingerprinting

Multiplex PCR reaction was performed successfully, giving a genetic profile for each of the 320 Arabica samples, as shown in Fig. 3. Alleles were used to calculate genetic distances and to design a cladogram showing genetic relationships between the samples. These varieties, whose origin is known over confidence, constitute a "genetic bank" that can be used to compare samples with unknown origin.

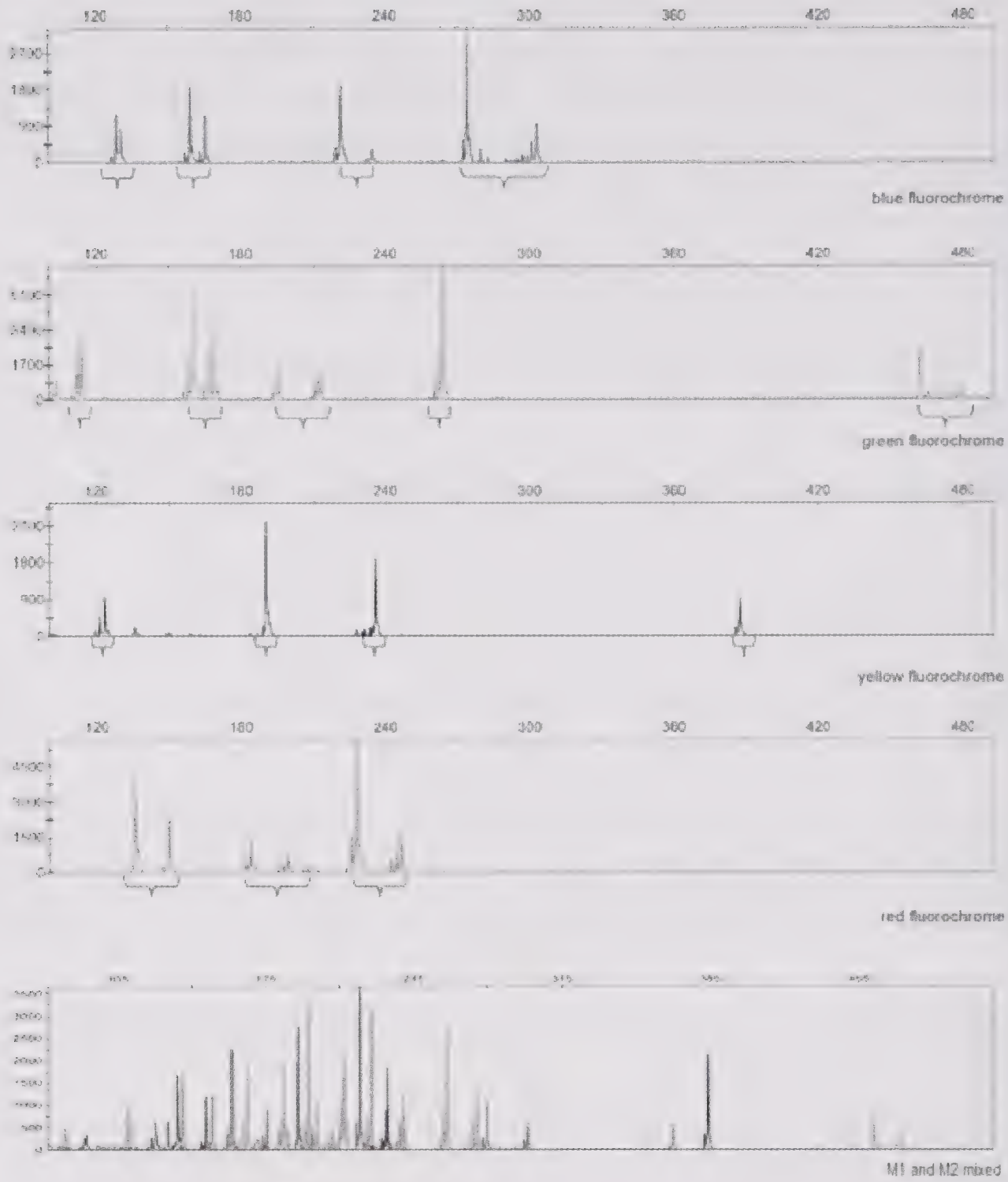


Fig. 3 – Example of sample amplification by Multiplex PCR.

4 Conclusions

All methods described in this paper can be successfully applied on green and roasted coffee beans with DNA extraction and analysis protocols set up in the Laboratory of Genetics of the University of Trieste. The main goals of these analyses aim towards food traceability. The authenticity of Arabica, Robusta or blends of the two species, is an important topic for producers and customers. Several molecular markers are available to establish the origin of coffee varieties for scientific aims, but none of them was used and validated for commercial purpose so far. The possible applications are many: analysis of a coffee stock proposed to a wholesaler, commercial coffee analysis to protect the dealer from unfair competition (food traceability). The analysis can be also used to determine the variety in a gourmet coffee lot.

References

[1] F. Carrera, M. Leon-Camacho, F. Pablos and A. G. Gonzalez, "Authentication of green coffee varieties according to their sterolic profile" *Anal. Chim. Acta,* vol. 370, pp.131-139, 1998.

[2] M. J. Martin, F. Pablos and A. G. Gonzalez, "Discrimination between arabica and robusta green coffee varieties according to their chemical composition" *Talanta,* vol. 46, pp. 1259-1264, 1998.

[3] M. J. Martin, F. Pablos, A. G. Gonzalez, M. S. Valdenebro and M. Leòn-Chamacho, "Fatty acid profiles as discriminant parameters for coffee varieties differentiation" *Talanta,* vol. 54, pp. 291–297, 2001.

[4] G. N. Jham, J. K. Winkler, M. A. Berhow and S. F. Vaughn, "γ-Tocopherol as a marker of Brazilian coffee (*Coffea arabica L.*) adulteration by corn" *J. Agric. Food Chem.,* vol. 55, pp. 5995-5999, 2007.

[5] S. Spaniolas, S.T. May, M. J. Bennett and G. A. Tucker, "Authentication of Coffee by Means of PCR-RFLP Analysis and Lab-on-a-Chip Capillary Electrophoresis" *J. Agric. Food Chem,* vol. 54(20), pp. 7466-7470, 2006.

[6] B. K. Chakravarthi and R. Naravaneni, "SSR marker based DNA fingerprinting and diversity study in rice (*Oryza sativa*. L)" *African J. Biotech.,* vol. 5(9), pp. 684-688, 2006.

[7] V. Ashkenazi, E. Chani, U. Lavi, D. Levy, J. Hillel and R. E. Veilleux, "Development of microsatellite markers in potato and their use in phylogenetic and fingerprinting analyses" *Genome,* vol. 44, pp. 50–62, 2001.

[8] M. M. Manifesto, A. R. Schlatter, H. E. Hopp, E. Y. Suárez and J. Dubcovsky, "Quantitative Evaluation of Genetic Diversity in Wheat Germplasm Using Molecular Markers" *Crop Science,* vol. 41, pp. 682-690, 2001.

[9] "Method for the discrimination between the varieties of Coffea arabica based on polymorphisms of nuclear DNA" Patent n. *PD2008A000336*.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 315-322.
ISBN 978-88-8303-295-0. EUT, 2010.

Mislabelling in megrims: implications for conservation

Victor Crego-Prieto, Daniel Campo, Juliana Perez,
Eva Garcia-Vazquez

Abstract — Mislabelling of fish catch and commercial seafood products is relatively frequent worldwide and can affect fisheries management exploitation when stock estimates are based on landings. In this study we have analyzed genetically 239 commercial lots of two morphologically similar species of megrims (genus *Lepidorhombus*) that are caught together in mixed fisheries. A high proportion of mislabelling was detected, suggesting enormous underreported exploitation of one of the species, which can be endangered if the problem persists. These results highlight the urgency of applying currently available species-specific molecular tools in fisheries sciences for preventing biodiversity losses in exploited species.

Index Terms — exploitation, genetic identification, megrims, mixed fisheries, species-specific markers.

1 Introduction

Industrialized fisheries typically reduce community biomass by 80% within 15 years of exploitation, and, as a consequence, large predatory fish biomass today is only about 10% of pre-industrial levels [1]. Depletion of fish stocks is due to many different factors, some of them anthropogenic. For example, due to factors ranging from climate change [2] to pollution or overfishing, exploited natural populations are in decline in many marine areas. Possible solutions for environmental challenges fall out of the scope of this study. Solutions for overfishing, however, exist and are relatively simple, although they may have a short-term socioeconomic cost for the fishery sector. Some solutions really work, as demonstrated by the recovery of Atlantic herring after its depletion in the late 70s and further implementation of protective measures [3]. Restricted fisheries effort and protection of spawning areas and juveniles are some of the

V. Crego-Prieto is with the Department of Functional Biology, University of Oviedo,C/ Julian Claveria s/n, 33003-Oviedo, Spain. E-mail: victor.crego.prieto@gmail.com.
D. Campo is with the Department of Molecular and Computational Biology, University of Southern California, 1050 Childs Way, RRI. Los Angeles (CA), 90089, USA. E-mail: dcampo@usc.edu.
J. Perez is with the Department of Natural Resources Conservation, Holdsworth Hall, Amherst MA 01003, USA.
E. Garcia-Vazquez is with the Department of Functional Biology, University of Oviedo,C/ Julian Claveria s/n, 33003-Oviedo, Spain. E-mail: egv@uniovi.es.

possible approaches for allowing natural stocks to recover. However, population estimation techniques are not exact, leading to inaccurate estimates of stock size. Eggs and larvae of different species with overlapping spawning areas are often morphologically similar, and methods of species identification in addition to visual identification are needed for accurate stock assessment [4], [5], [6].

The same problem exists when estimates of fishery effort are based on reported catch data. In some cases, such as for sharks in Hong Kong markets, high concordance between trade and specific names may allow the use of market records for monitoring species-specific trends in trade and exploitation rates [7], [8]. This method, however, cannot be generalized. Sometimes the adults of two species caught simultaneously, for example in trawl fisheries, are so similar that it is difficult to identify them so that mislabelling may occur, as shown for example in hakes [9]. Once mislabelled at landing, the error persists along the entire seafood chain to the consumer, who buys a marketed product which does not correspond to the species marked on the label.

DNA variants can be revealed employing many different techniques. Here it is almost impossible to describe all of them in detail, but some specific examples of molecular techniques used for revealing species-specific variations in fish are listed in Tab 1. They could be useful to fisheries if assessment of trade fish products is used to estimate stock exploitation.

The aim of this study was to analyze in detail a case where application of species-specific markers to fisheries science seems necessary and likely urgent. Species misidentification was detected from landings to commercial products, suggesting underreported exploitation of megrim species (genus *Lepidorhombus*), whose exploitation rates are largely based on catch reports. We also assessed the possible consequences of these likely inadvertent errors for long-term sustainability of fish stocks.

2 Material and methods

2.1 Mixed-fisheries case studies

We have focused our study on two morphologically similar fish species that are caught in different areas of the Atlantic Ocean. In Europe, the two species of megrim, *Lepidorhombus whiffiagonis* (megrim) and *L. boscii* (four-spotted megrim), are flatfishes of the Scophthalmidae family (Pleuronectiformes) having overlapping distributions (Fig. 1).

As with other species, they are caught together in trawl fisheries. Most landings correspond to Spain, followed by the UK, the two countries having together approximately 70% of European catches. In the period 2000-2004, catches of 58,180 tons of *L. whiffiagonis* were reported (FAO catch statistics, available at http://www.fao.org/fishery/statistics/global-capture-production) compared to only 40,187 tons of *L. boscii* (40.85 % of megrim catches). Little is known about population structure and/or population size of megrims, although the existence of a separate stock of *L. whiffiagonis* in the Mediterranean sea has been demonstrated employing genetic markers [18] and differences in growth

among distributional areas have been described [19], [20].

2.2 Samples analyzed

Reference samples for each species (Tab. 2) were obtained in the context of research cruises for the European Project MARINEGGS. The specimens were obtained from at least four different locations covering roughly the Atlantic distribution range of each species and identified by local experts in fish taxonomy. A piece of gill or muscle tissue (about 3 g) was taken from each specimen and stored in absolute ethanol. The reference samples are deposited in the laboratory of the research team at the University of Oviedo (Spain).

Spain was chosen for the survey because it is the top country in megrim fisheries (43% of total landings, FAO catch statistics 2008). Marketed products of both megrim species, labelled with species names, were directly purchased from landings in Asturias (North Spain) in 2004. A total of 239 landings were analyzed. A piece of tissue (approx. 3 g) was taken from each sample and stored in absolute ethanol until analysis.

2.3 DNA analyses

DNA extraction was carried out employing the resin Chelex [21]. For identification of megrims we used differences in sequence length of the PCR-amplified fragment of a conserved locus, the 5S rDNA coding for small ribosomal RNA (5S rRNA), as the species-specific marker. The locus is composed of the coding sequence, typically 120 base pairs (bp) long and highly conserved among species, and the non transcribed spacer (NTS) which can differ in sequence and length among closely related species. PCR amplification of the 5SrDNA locus was carried out in a GeneAmp PCR system 9700 (Applied Biosystems), employing the primers designed by Pendas et al. [22], in a total volume of 20 µl containing 0.5 µl GoTaq Polymerase at 5U/ml (Promega), 2µl of 10x Buffer, 2µl of 25 mM $MgCl_2$, 2µl of dNTPs, 100 pmol of each primer and approximately 5 ng of genomic DNA. PCR amplification conditions were: initial denaturation at 95°C for 5 min, then 35 cycles of denaturation at 95°C for 20 s, annealing at 65° for 20 s and extension at 72°C for 30 sec, and a final extension at 72°C for 20 min. When agarose methodology was employed, products were run in 2.5% agarose gels at 100 V and visualized by staining with 2 µl ethidium bromide (10 mg/ml). The size of the amplified fragments was estimated by comparison with a standard 100 bp DNA marker (Promega). In the Genetic Analyzer (Sequencing Unit, University of Oviedo), fragment sizes were also directly visualized in a chromatogram employing the GeneScan 3.7 Analysis Software (Applied Biosystems).

3 Results

3.1 Species-specificity of the markers assayed

Species-specificity of the marker was confirmed for the two species studied. All the individuals morphologically identified as belonging to a given species yielded the same genetic pattern. The 5S rRNA locus amplification yielded two DNA fragments of 435 and 217 base pairs (bp) for *Lepidorhombus whiffiagonis* and two main fragments 331 and 233 bp long (plus some secondary heavier shorter fragments) for *L. boscii* (Fig. 2, amplification fragments visualized in an agarose gel).

3.2 *Mislabelling in landings*

A high level of mislabelling was found in the 239 commercial landings analyzed. Although declared landings were 60% *L. whiffiagonis* and 40% *L. boscii*, the actual proportion of each species was 49 and 51% for *L. whiffiagonis* and *L. boscii,* respectively (Tab. 3).

4 Discussion

The 5S rDNA can be considered a good species-specific marker. It has already been used for example in fishes like the Genus *Leporinus* [23], the Sciaenidae family [24] or in shark species [25], and also in many other taxa. The patterns obtained here for both megrims are in concordance with patterns described for this species by Garcia-Vazquez et al. [26]. Mislabelling in these megrim species is likely accidental, as they are morphologically similar and often difficult to separate by visual inspection. The trade price is the same for the two species, therefore intentional mislabelling for purposes of commercial fraud cannot explain the detected differences between declared and real commercialized species. Although inadvertent, this type of mislabelling could however produce serious errors in fisheries assessments. If we assume that the individuals analyzed are representative of landings, the divergence between declared and actual catches would be thousands of tons of megrims (Fig. 3). Figures corresponding to estimated “actual” catch can be obtained based on species content in the commercial landings analysed (in percent), multiplied by the total catch (in tons) of each species.

Another point to consider is the direction of mislabelling, which was deviated incrementing the catch data of *L. whiffiagonis* in a high percentage, and so decreasing the catch data corresponding to *L. boscii*. Underreported exploitation of a species leads to overexploitation and, in the long term, to exhaustion of stocks, fisheries decline and eventual extinction of the overexploited species [27]. For purposes of fisheries management, these data should clearly be taken into account.

Stock sizes are not estimated separately for this two species in annual surveys in their respective area of occurrence. The two *Lepidorhombus* are not

genetically distinguished in routine plankton surveys, although there are recent studies describing species-specific markers for these species [5] that clearly demonstrate that visual identification is not accurate. *L. whiffiagonis* was the only megrim species identified in Bay of Biscay plankton samples [28], and was also the megrim species confounded with hake eggs in other plankton surveys [5]. Absence to date of genetically analyzed *L. boscii* in plankton samples could be interpreted as a signal of its scarcity, but those studies were based on a limited number of samples and cannot be taken as an indicator of real abundance of that species.

Genetic identification of specimens in landings is even more important for species like those studied in this work, whose production in aquaculture is not forecasted at short-term. As demersal species, their cultivation is not easy. For megrims, cultivation assays have not been carried out as far as we know. Thus, although aquaculture seems to be a solution for obtaining seafood protein at a global scale, as for other marine species [29], production of megrim at commercial scale will likely rely on extractive fisheries in the forthcoming years. Application of species-specific markers to fisheries science seems necessary and likely urgent, and stock evaluation based on catch records will require application of genetic markers for improving its utility for sustainable exploitation of these valuable marine species.

5 Conclusion

DNA analysis revealed high percentage of mislabelling in megrim landings. These results suggest underreported exploitation of four-spotted megrim *L. boscii*, a species whose exploitation rates are largely based on catch reports and which could become endangered if the problem persists. We highlight the urgency of applying currently available species-specific molecular tools in fisheries sciences.

Acknowledgements

We thank Paula Alvarez (AZTI, Spain), Francisco Sanchez (IEO, Santander, Spain) and Placida Lopes (IPIMAR, Portugal) for providing megrim samples. This study was supported by the FICYT project IB09-0023 (Asturias, Spain). Ivan Gonzalez Pola provided help with laboratory analyses. Eva Garcia-Vazquez was a Grantee from the Spanish Ministry of Research and Innovation (PR2008-0239) in 2008.

References

[1] R. A. Myers and B. Worm, "Rapid worldwide depletion of predatory fish communities". *Nature,* vol. 423, pp. 280-283, 2003.

[2] C. M. O'Brien, C. J. Fox, B. Planque et al., "Fisheries: climate variability and North Sea cod". *Nature,* vol. 404, p. 142, 2000.

[3] J. A. Hutchings, "Collapse and recovery of marine fishes". *Nature* 406, pp. 882-885, 2000.

[4] C. J. Fox, M. I. Taylor, R. Pereyra et al., "TaqMan DNA technology confirms likely overestimation of cod (*Gadus morhua L.*) egg abundance in the Irish Sea: implications for the assessment of the cod stock and mapping of spawning areas using egg-based methods". *Molecular Ecology,* vol. 14(3), pp. 879–884, 2005.

[5] J. Perez, P. Alvarez, J. L. Martinez et al., "Genetic identification of hake and megrim eggs in formaldehyde-fixed plankton samples". *ICES Journal of Marine Science* 62, 908-914, 2005a.

[6] M. Kochzius, M. Nölte, H. Weber et al., "DNA Microarrays for Identifying Fishes". *Marine Biotechnology,* vol. 10(2), pp. 207-217, 2008.

[7] D. L. Abercrombie, S. C. Clarke and M. S. Shivji, "Global-scale genetic identification of hammerhead sharks: Application to assessment of the international fin trade and law enforcement". *Conservation Genetics,* vol. 6, pp. 775-788, 2005.

[8] S. C. Clarke, M. K. McAllister, E. J. Milner-Gulland et al., "Global estimates of shark catches using trade records from commercial markets". *Ecology Letters,* vol. 9, pp. 1115-1126, 2006.

[9] G. Machado-Schiaffino, J. L. Martinez and E. Garcia-Vazquez, "Detection of mislabeling in hake seafood employing mtSNPs-based methodology with identification of eleven hake species of the genus Merluccius". *Journal of Agriculture and Food Chemistry,* vol. 56(13), pp. 5091-5095, 2008.

[10] F. Teletchea, "Molecular identification of fish species; reassessment and possible applications". *Reviews in Fish Biology and Fisheries* (on line first DOI 10.1007 1160-009-9107-4), 2009.

[11] D. Blohm, F. Bonhomme, G. Carvalho et al., "Assessment of tools for identifying the genetic origin of fish and monitoring their occurrence in the wild". In: T. Svåsand, D. Crosetti, E. Garcia-Vazquez and E. Verspoor (eds.), *Genetic impact of aquaculture activities on native populations* (Genimpact final scientific report, EU contract n. RICA-CT-2005-022802). http://genimpact.imr.no/, pp. 128-134, 2007. Accessed 12 March 2009.

[12] M. M. Ferguson and R. G. Danzmann, "Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting?" *Canadian Journal of Fisheries and Aquatic Sciences,* vol. 55(7), 1553-1563, 1998.

[13] Z. J. Liu and J. F. Cordes, "DNA marker technologies and their applications in aquaculture genetics". *Aquaculture,* vol. 238, pp. 1-37, 2004.

[14] R. S. Rasmussen and M. T. Morrissey, "DNA-based methods for the identification of commercial fish and seafood species". *Comprehensive Reviews in Food Science and Food Safety,* vol. 7, pp. 280-295, 2008.

[15] F. Aranishi, T. Okimoto and S. Izumi, "Identification of gadoid species (Pisces, Gadidae) by PCR-RFLP analysis". *Journal of Applied Genetics,* vol. 46(1), pp. 69-73, 2005.

[16] M. I. Taylor, C. Fox, I. Rico and C. Rico, "Species-specific TaqMan probes for simultaneous identification of (*Gadus morhua L.*), haddock (*Melanogrammus aeglefinus L.*) and whiting (*Merlangius merlangus L.*)". *Molecular Ecology Notes,* vol. 2(4), pp. 599-601, 2002.

[17] J. E. Magnussen, E. K. Pikitch, S. C. Clarke et al., "Genetic tracking of basking shark products in international trade". *Animal Conservation,* vol. 10(2), pp. 199-207, 2007.

[18] E. Garcia-Vazquez, J. I. Izquierdo and J. Perez, "Genetic variation at ribosomal genes supports the existence of two different European subspecies in the megrim *Lepidorhombus whiffiagonis*". *Journal of Sea Research,* vol. 56, pp. 59-64, 2006a.

[19] J. Landa and C. Piñeiro, "Megrim (*Lepidorhombus whiffiagonis*) growth in the North-eastern Atlantic based on back-calculation of otolith rings". *ICES Journal of Marine Science,* vol. 57(4), pp. 1077-1090, 2000.

[20] J. Landa, N. Perez and C. Piñeiro, "Growth patterns of the four spot megrim in the northeast Atlantic". *Fisheries Research,* vol. 55, pp. 141-152, 2002.

[21] A. Estoup, C. R. Largiader, E. Perrot. et al., "Rapid one-tube DNA extraction protocol for reliable PCR detection of fish polymorphic markers and transgenes". *Molecular Marine Biology and Biotechnology,* vol. 5(4), pp. 295-298, 1996.

[22] A. M. Pendas, P. Moran, J. L. Martinez and E. Garcia-Vazquez, "Applications of 5S rDNA in Atlantic salmon, brown trout, and in Atlantic salmon x brown trout hybrid identification". *Molecular Ecology,* vol. 4, pp. 275-276, 1995.

[23] I. A. Ferreira, C. Oliveira, P. C. Venere, P. M. Galetti Jr and C. Martins, "5S rDNA variation and its phylogenetic inference in the genus *Leporinus* (*Characiformes: Anostomidae*)". *Genetica,*

vol. 129(3), pp. 253-257, 2006.

[24] F. A. Alves-Costa, C. Martins, F. Del Campos de Matos, F. Foresti, C. Oliveira and A. P. Wasko, "5S rDNA characterization in twelve Sciaenidae fish species (Teleostei, Perciformes): depicting gene diversity and molecular markers". *Genetic Molecular Biology,* vol. 31(1) Suppl. 0, 2008.

[25] D. Pinhal, O. B. F. Gadig, A. P. Wasko, C. Oliveira, E. Ron, F. Foresti and C. Martins, "Discrimination of shark species by simple PCR of 5S rDNA repeats". *Genetic Molecular Biology,* vol. 31(1), pp. 361-365, 2008.

[26] E. Garcia-Vazquez, J. I. Izquierdo and J. Perez, Genetic variation at ribosomal genes supports the existence of two different European subspecies in the megrim *Lepidorhombus whiffiagonis*". *Journal of Sea Research,* vol. 56, pp. 59-64, 2006b.

[27] D. J. Agnew, J. Pearce, G. Pramod, T. Peatman, R. Watson, J. R. Beddington and T. J. Pitcher, "Estimating the Worldwide Extent of Illegal Fishing". PLoS One www.plosone.org, vol. 4(2), pp. 1-8, 2009.

[28] E. Garcia-Vazquez, P. Alvarez, P. Lopes. et al., "PCR-SSCP of the 16S rRNA gene, a simple methodology for species identification of fish eggs and larvae". *Scientia Marina,* vol. 70 (Suppl. 2), pp. 13-21, 2006c.

[29] R. Goldburg and R. Naylor, "Future seascapes, fishing, and fish farming". *Frontiers in Ecology and the Environment,* vol. 3, pp. 21-29, 2005.

[30] F. Aranishi, T. Okimoto and S. Izumi, "Identification of gadoid species (Pisces, Gadidae) by PCR-RFLP analysis". *Journal of Applied Genetics,* vol. 46(1), pp. 69-73, 2005.

[31] L. Asensio, I. González, M. A. Rodríguez, B. Mayoral, I. López-Calleja, P. E. Hernández, T. García and R. Martín, "Identification of grouper (*Epinephelus guaza*), wreck fish (*Polyprion americanus*), and Nile perch (*Lates niloticus*) fillets by polyclonal antibody-based enzyme-linked immunosorbent assay". *Journal of Agriculture and Food Chemistry,* vol. 51(5), pp. 1169-1172, 2003.

[32] L. Asensio, I. González, M. A. Pavón, T. García and R. Martín, "An indirect ELISA and a PCR technique for the detection of Grouper (*Epinephelus marginatus)* mislabeling". *Food Additives and Contamination,* vol. 25(6), pp. 677-683, 2008.

[33] E. Carrera, T. García, A. Céspedes et al., "Differentiation of smoked Salmo salar, Oncorhynchus mykiss and Brama raii using the nuclear marker 5S rDNA". *International Journal of Food Science and Technology,* vol. 35, pp. 401-406, 2000.

[34] A. G. F. Castillo, J. L. Martinez and E. Garcia-Vazquez, "Identification of Atlantic hake species by a simple PCR-based methodology employing microsatellite loci". *Journal of Food Protection,* vol. 66(11), pp. 2130-2134, 2003.

[35] M. Carrera, B. Cañas, C. Piñeiro, J. Vázquez and J. M. Gallardo, "Identification of commercial hake and grenadier species by proteomic analysis of the parvalbumin fraction". *Proteomics ,* vol. 6(19), pp. 5278-5287, 2006.

[36] M. Carrera, B. Cañas, C. Piñeiro, J. Vázquez and J. M. Gallardo, "De novo mass spectrometry sequencing and characterization of species-specific peptides from nucleoside diphosphate kinase B for the classification of commercial fish species belonging to the family Merlucciidae". *Journal of Proteome Research,* vol. 6(8), pp. 3070-3080, 2007.

[37] M. J. Chapela, A. Sánchez, M. I. Suárez, R. I. Pérez-Martín and C. G. Sotelo, "A rapid methodology for screening hake species (*Merluccius spp.*) by single-stranded conformation polymorphism analysis". *Journal of Agriculture and Food Chemistry,* vol. 55(17), pp. 6903-6909, 2007.

[38] Z. Hubalkova, P. Kralik, J. Kasalova and E. Rencova, "Identification of gadoid species in fish meat by polymerase chain reaction (PCR) on genomic DNA". *Journal of Agriculture and Food Chemistry,* vol. 56(10), pp. 3454-3459, 2008.

[39] D. F. Hwang, H. C. Jen, Y. W. Hsieh and C. Y. Shiau, "Applying DNA techniques to the identification of the species of dressed toasted eel products". *Journal of Agriculture and Food Chemistry,* vol. 52(19), pp. 5972-5977, 2004.

[40] R. Murgia, G. Tola, S. N. Archer, S. Vallerga and J. Hirano, "Genetic identification of grey mullet species (Mugilidae) by analysis of mitochondrial DNA sequence: application to identify the origin of processed ovary products (bottarga)". *Marine Biotechnology,* vol. 4(2), pp. 119-126, 2002.

[41] J. Perez and E. Garcia-Vazquez, Genetic identification of nine hake species for detection of commercial fraud. *Journal of Food Protection,* vol. 67, 2792-2796, 2004.

[42] M. Perez, J. M. Vieites and P. Presa, "ITS1-rDNA-based methodology to identify world-wide hake species of the Genus *Merluccius*". *Journal of Agriculture and Food Chemistry,* vol. 53(13), pp. 5239-5247, 2005b.

[43] C. Piñeiro, J. Barros-Velázquez, R. I. Pérez-Martín et al., "Development of a sodium dodecyl sulfate-polyacrylamide gel electrophoresis reference method for the analysis and identification of fish species in raw and heat-processed samples: a collaborative study". *Electrophoresis,* vol. 20(7), pp. 1425-1432, 1999.

[44] C. Piñeiro, J. Vázquez, A. I. Marina, J. Barros-Velázquez and J. M. Gallardo, "Characterization and partial sequencing of species-specific sarcoplasmic polypeptides from commercial hake species by mass spectrometry following two-dimensional electrophoresis". *Electrophoresis,* vol. 22(8), pp. 1545-1552, 2001.

[45] I. Rodushkin, T. Bergman, G. Douglas, E. Engström, D. Sörlin and D. C. Baxter, "Authentication of Kalix (N.E. Sweden) vendace caviar using inductively coupled plasma-based analytical techniques: evaluation of different approaches". *Analytica Chimica Acta,* vol. 583(2), pp. 310-318, 2007.

[46] P. Sebastio, P. Zanelli and T. M. Neri, "Identification of anchovy (*Engraulis encrasicholus L.*) and gilt sardine (*Sardinella aurita*) by polymerase chain reaction, sequence of their mitochondrial cytochrome b gene, and restriction analysis of polymerase chain reaction products in semipreserves". *Journal of Agriculture and Food Chemistry,* vol. 49(3), pp. 1194-1199, 2001.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 323-326.
ISBN 978-88-8303-295-0. EUT, 2010.

Seeds in subtribe Orchidinae (Orchidaceae): the best morphological tool to support molecular analyses

Roberto Gamarra, Emma Ortúñez, Ernesto Sanz,
Iris Esparza, Pablo Galán

Abstract — Seeds of the genera belonging to the subtribe Orchidinae (Orchidaceae) have been studied using scanning electron microscopy. Qualitative data concerning the general morphology of the seed and testa cells, and ornamentation of periclinal and anticlinal walls, are characters that allow to recognize genera and taxa at infrageneric levels. The results show that seed micromorphology is the best tool to support the molecular analyses published in recent years for this taxonomic group.

Index Terms — qualitative data, seed micromorphology, SEM, subtribe Orchidinae.

1 Introduction

The family Orchidaceae is probably the largest among flowering plants, with 25.158 species [1]. The systematics has undergone many changes along the last few decades. The latter taxonomic proposals were published by Dressler [2] and Szlachetko [3]. In the subfamily Orchidoideae, Dressler [2] divided the tribe Orchideae into two subtribes: Orchidinae with 34 genera and 370 species, and Habenariinae with 23 genera and 930 species. Within this tribe, Philip Cribb [4] recognizes 62 genera and near 1800 species. This is the principal orchid group of the north temperate area, also with considerable diversity in Africa, South Asia and Australia; only the genus *Habenaria* occurs in South America.

Flower morphology (lip, anther, stigma, rostellum) and vegetative characters (habitus, inflorescence, tuberoids) have been used to elucidate the systematics of this tribe. According to Dressler [2], the morphology of tuberoids is essential

R. Gamarra, E. Ortúñez, E. Sanz and I. Esparza are with the Departamento de Biología, Universidad Autónoma de Madrid, C/ Darwin, 2, E-28049 Madrid, Spain. E-mail: roberto.gamarra@uam.es. P. Galán is with the Departamento de Producción Vegetal: Botánica y Protección Vegetal, E.U.I.T. Forestal, Universidad Politécnica, E-28040 Madrid, Spain, E-mail: pablo.galan@uam.es.

to recognise the genera in the subtribe Orchidinae, while in Habenariinae the stigmas are more important. Szlachetko [3] proposes 6 subtribes (Orchidinae, Herminiinae, Bartholininae, Androcorytinae, Platantherinae, Habenariinae) based on stigma and rostellum morphology.

Recently, molecular analyses have changed the taxonomy of several genera and species in this tribe [5], [6], e.g., the monotypic genus *Coeloglossum* Hartm. is integrated into *Dactylorhiza* Neck. ex Nevski, while some species attributed to the genus *Orchis* L. are included into *Anacamptis* Rich. and *Neotinea* Rchb. fil.

Beer [7] published the first study about the seed morphology in Orchidaceae. In his book, the figures show the great diversity in genera belonging to different subfamilies.

Clifford & Smith [8] proposed the first methodology to analyze the testa morphology studying 49 species of the subfamily Epidendroideae. They showed a strong correlation between qualitative characters of seeds and taxa above genus level.

Later, Barthlott [9] studied 58 genera using SEM, demonstrating the great diagnostic value of seed morphology and its phylogenetic significance, principally at tribe and subtribe level. He also indicated that this is a useful taxonomic tool to recognise the genera. In 1979, Joseph Arditti and colleagues established the methodology for quantitative analyses, related to the sizes and volumes of seeds and embryos [10]. Several authors published different papers about seed morphology in the family Orchidaceae [11], [12], [13]. All these papers show the great diagnostic value of the qualitative and quantitative characters of the seed, to approach phylogenetic studies in this family. Arditti & Al-Ghani [14] published an overview of previous works, and they conclude the importance to continue this research.

More recently. Krishna Swamy et al. [15] described seeds of the genus *Cymbidium* using SEM and morphometric data, and Tsutsumi et al. [16] compared the phylogenetic proposal using molecular markers with seed morphology of the japanese species of the genus *Liparis.*

Different authors have obtained similar results in other flowering plants, as in the genus *Phyllocladus* (Phyllocladaceae) [17], in the tribe Massonieae (Hyacinthaceae) [18], in the genus Veronica (Plantaginaceae) [19], and in the genus Moehringia (Caryophyllaceae) [20]. These papers show that seed morphology is an important tool to elucidate the taxonomy of these genera.

Since 2003, our research group initiated a study of seed micromorphology of iberian orchids using SEM and the methodology proposed by former authors. We have published data about genera of subfamilies Cypripedioideae [21] and Orchidoideae [22], [23], some of them previously presented in the 17th International Botanical Congress, held in Wien in 2005. Our studies show that qualitative and quantitative characters strongly support the results using molecular markers. These characters are of good diagnostic value to recognize many taxa, principally above the species level.

Presently, the main aim of our research is the study of the seed micromorphology in all groups within the genera of the tribe Orchideae.

2 Results

In the studied genera, the seed morphology varies from elongate fusiform (*Dactylorhiza, Platanthera, Neotinea, Anacamptis)*, to shortly fusiform and almost ovoid (*Gymnadenia, Pseudorchis, Amitostigma)*. Generally, in the elongate fusiform, the medial cells are longer than apical and basal cells; in the other morphologycal type, are similar or slightly longer.

The apical pole mainly consist of short and polygonal cells. Only the genus *Platanthera* finished in a truncated cell.

The chalazal end is opened, with short and polygonal cells. Exclusively, in the genus *Ophrys*, a distinct asymmetry is showed.

In several genera (*Orchis, Pseudorchis, Gymnadenia)*, the periclinal walls are unsculptured. However, many genera present a type of ornamentation, with prominent and spaced ridges *(Ophrys)*, to slight ondulations *(Serapias, Amerorchis, Comperia)*. The distribution of ridges and ondulations varies from lax *(Platanthera, Himantoglossum)* to dense *(Serapias, Steveniella, Comperia)*, and from transversal *(Neotinea, Steveniella*) to oblique *(Anacamptis, Himantoglossum, Aorchis)*. Only the genus *Dactylorhiza* shows a great variation, from unsculptured periclinal walls *(D. incarnata* group) to different types of ornamentation *(D. maculata* group, *D. majalis* group).

The morphology of the anticlinal walls varies from straight *(Platanthera, Himantoglossum, Aorchis*) to undulate *(Gymnadenia, Orchis* p.p., *Anacamptis* p.p.) The las type is more typical in the cells of apical pole. Also, a distinct type of lamella can be found in these walls *(Ophrys).*

3 Conclusion

The morphological study, including qualitative and quantitative characters, of the seed coat in the genera of the subtribe Orchidinae has showed that each genus and each subgroup within this, have its own morphological type. Each type fully agrees with the clades obtained in the recently published molecular analyses [5], [6]. For example, it is consistent with the inclusion of the genus *Coeloglossum* with the species of the *Dactylorhiza incarnata* group, *Nigritella* into *Gymnadenia* or *Barlia* into *Himantoglossum*. Also, within genera with many species as *Anacamptis, Orchis* or *Dactylorhiza*, each clade is consistent with each morphological seed type. Likewise, it also supports the monophyly of genera such as *Ophrys* and *Serapias*.

Acknowledgement

We are much indebted to Esperanza Salvador, Enrique and Isidoro, the technical staff in the SEM laboratory at SIDI-UAM. To Jeff Wood and Mauricio Velayos, the curators of Royal Botanic Gardens at Kew and Madrid, respectively. We are also grateful to all orchidologists who sent us seeds of different genera.

References

[1] P. Cribb and R. Govaerts, "Just how many orchids are there", *Proc. 18th World Orchid Conference*, pp. 161-172, 2005.

[2] R. Dressler, *Phylogeny and classification of the orchid family*, Cambridge, Cambridge University Press, 1993.

[3] D. Szlachetko, "Systema Orchidalium", *Fragm. Florist. Geobot.*, Suppl. 3, pp. 1-152, 1995.

[4] A. M. Pridgeon, P. Cribb, M. W. Chase and F.N. Rasmussen, "*Genera Orchidacearum. 2. Orchidoideae (Part 1)"*, Oxford, Oxford University Press, 2001.

[5] A. M. Pridgeon, R. M. Bateman, A. V. Cox, J. R. Hapeman and M. W. Chase, "Phylogenetics of subtribe Orchidinae (Orchidoideae, Orchidaceae) based on nuclear ITS sequences. 1. Intergeneric relationships and polyphyly of *Orchis* sensu lato". *Lindleyana*, vol. 12, pp. 89-109, 1997.

[6] R. M. Bateman, P. M. Hollingsworth, J. Preston, L. Yi-Bo, A. M. Pridgeon and M. W. Chase, "Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae)", *Bot. J. Linn. Soc.*, vol. 142, pp. 1-40, 2003.

[7] J. G. Beer, *Beiträge zur Morphologie und Biologie der familie der orchideen*, Vienna, Druck und Verlag von Carl Gerold's Sohn, 1863.

[8] H. T. Clifford and W. K. Smith, "Seed morphology and classification of Orchidaceae", *Phytomorphology*, vol. 19, pp. 133-139, 1969.

[9] W. Barthlott, "Morphologie der Samen von Orchideen im Hinblick auf taxonomische und funktionelle Aspekte", *Proc. 8th World Orchid Conference,* pp. 444-455, 1976.

[10] J. Arditti, J.D. Michaud and P. L. Healey, "Morphometry of orchid seeds. I. *Paphiopedilum* and native California and related species of *Cypripedium*", *Amer. J. Bot.*, vol. 66, pp. 1128-1137, 1979.

[11] H. Tohda, "Seed morphology in Orchidaceae I. *Dactylorchis, Orchis, Ponerorchis, Chondradenia* and *Galeorchis"*, *Sci. Rep. Tohoku Univ., 4th ser., Biology*, vol. 38, pp. 253-268, 1983.

[12] H. Kurzweil, "Seed morphology in Southern African Orchidoideae (Orchidaceae)", *Pl. Syst. Evol.*, vol.185, pp. 229-247, 1993.

[13] M. Molvray and P. J. Kores, "Character analysis of the seed coat in Spiranthoideae and Orchidoideae, with special reference to the Diurideae (Orchidaceae)". *Amer. J. Bot.*, vol. 82, pp. 1443-1454, 1995.

[14] J. Arditti and A. K. Al-Ghani, "Numerical and physical properties of orchid seeds and their biological implications", *New Phytologist,* vol. 145, pp. 367-421, 2000.

[15] K. Krishna Swamy, H. N. Krishna Kumar, T. M. Ramakrishna and S. N. Ramaswamy, "Studies on seed morphometry of epiphytic orchids from Western Ghats of Karnataka", *Taiwania*, vol. 49, pp. 124-140, 2004.

[16] C. Tsutsumi, T. Yukawa, N. Lee, C. Lee and M. Kato, "Phylogeny and comparative seed morphology of epiphytic and terrestrial species of *Liparis* (Orchidaceae) in Japan", *J. Pl. Res.*, vol. 120, pp. 405-412, 2007.

[17] A. V. Bobrov, A. P. Melikian and E. Y. Yembaturova, "Seed morphology, anatomy and ultrastructure of *Phyllocladus* L. C. and A. Rich. ex Mirb. (Phyllocladaceae (Pilg.) Bessey) in connection with the generic system and Phylogeny", *Ann. Bot.*, vol. 83, pp. 601-618, 1999.

[18] W. Pfosser, W. Wetschnig, S. Ungar and G. Prenner, "Phylogenetic relationships among genera of Massonieae (Hyacinthaceae) inferred from plastid DNA and seed morphology". *J. Pl. Res.*, vol. 116, pp. 115-132, 2003.

[19] L. M. Muñoz-Centeno, D. C. Albach, J. A. Sánchez-Agudo and M. M. Martínez-Ortega, "Systematic significance of seed morphology in *Veronica* (Plantaginaceae): a phylogenetic perspective". *Ann. Bot.*, vol. 98, pp. 335-350, 2006.

[20] L. Minuto, S. Fior, E. Roccotiello and G. Casazza, "Seed morphology in *Moehringia* L. and its taxonomic significance in comparative studies within the Caryophyllaceae", *Pl. Syst. Evol.*, vol. 262, pp. 189-208, 2006.

[21] E. Ortúñez, E. Dorda, P. Galán and R. Gamarra, "Seed micromorphology in the iberian Orchidaceae. I. Subfamily Cypripedioideae", *Bocconea*, vol. 19, pp. 271-274, 2006.

[22] R. Gamarra, E. Dorda, A. Scrugli, P. Galán and E. Ortúñez, "Seed micromorphology in the genus *Neotinea* Rchb. f. (Orchidaceae, Orchidinae)" *Bot. J. Linn. Soc.*, vol. 153, pp. 133-140, 2007.

[23] R. Gamarra, P. Galán, I. Herrera and E. Ortúñez, "Seed micromorphology supports the splitting of *Limnorchis* from *Platanthera* (Orchidaceae)", *Nord. J. Bot.*, vol. 26, pp. 61-65, 2008.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 327-331.
ISBN 978-88-8303-295-0. EUT, 2010.

Lentils biodiversity: the characterization of two local landraces

Vincenzo Viscosi, Manuela Ialicicco, Mariapina Rocco, Dalila Trupiano, Simona Arena, Donato Chiatante, Andrea Scaloni, Gabriella Stefania Scippa

Abstract — A multi-disciplinary approach was used to characterize two autochthonous lentil landraces from Molise region (Central Italy). Different mature seed populations for each landrace were provided by the Molise Germoplasm Bank at the University of Molise (Pesche, Italy), and analyzed at the morphological and molecular (DNA and protein) levels. Nuclear ISSR markers were used to assess genetic differences, whereas phenotypic variability was detected by biochemical (proteomics) and morphological analyses. The genetic and phenotypic diversity of the two lentil landraces were well assessed in relation to their geographical provenance, supporting further studies to identify landrace markers.

Index Terms — ISSR markers, *Lens culinaris*, seed morphology, proteomics.

1 Introduction

Lens culinaris Medik. has been cultivated around the Mediterranean basin since at least the seventh century B.C. and its cultivation area expanded to Middle East, Ethiopia and the Indian Subcontinent ([1], [2]). Local landraces are characterized by high genetic variability and high adaptation to different environmental conditions evolving in adaptive gene complexes [3]. However, in the industrialized countries the cultivation of many different local landraces has progressively decreased, becoming at high risk of genetic erosion ([4], [5]). In this paper, the diversity of two lentil landraces is analyzed to define and quantify differences between groups of populations coming from two geographical areas of Molise region (Central Italy). In particular, this study aims

V. Viscosi, M. Ialicicco, D. Trupiano and G. S. Scippa are with the Department S.T.A.T. of University of Molise, Pesche, I-86090. E-mail: vincenzo.viscosi@unimol.it, manuela.ialicicco@unimol.it, dalila.trupiano@unimol.it, scippa@unimol.it.
M. Rocco is with the Department of Biological and Environmental Sciences, University of Sannio, Benevento, Italy. E-mail: rocco@unisannio.it.
S. Arena and A. Scaloni are with the Proteomics and Mass Spectrometry, ISPAAM, National Research Council, Naples, Italy.
D. Chiatante is with the Department S.C.A., University of Insubria, Como, Italy.

to deepen the knowledge about morphological, genetic and proteomic markers that differentiate local lentil landraces in relation to their provenance.

2 Material and Methods

Twelve lentil seed populations from Molise region (Central Italy) were studied in relation to their provenance: five from Conca Casale and seven from Capracotta. The total sample was analyzed by a multi-disciplinary approach, useful to characterize the phenotypical and genotypical traits of the two landraces. For the genetic analysis, nuclear ISSR markers were used as described in a previous paper [3], whereas the biochemical investigation was carried out on total seed proteins extracted according to the method of Rabilloud and resolved by 2DE [3]. For each population, gels were run in triplicate, and the mean value computed by software-assisted (PD-Quest) analysis was used to obtain a standardized matrix of the abundance (relative volume) of protein spots. One hundred seeds of each landrace population were used to measure eight morphological traits: area, perimeter, major axis length, minor axis length, and roundness, 100-seed weight, 100-seed volume and density (g/ml).

2.1 Statistical analysis

For molecular data, Principal Component Analysis was computed on Nei's genetic distance (1972) and the hierarchical partition of genetic variation among and within populations was obtained by means of the analysis of molecular variance (AMOVA). For biochemical and morphological data, a standardized matrix was subjected to univariate (ANOVA) and multivariate statistical analysis. Principal Component Analysis was computed on significant variables (detected by ANOVA) and the extracted Principal Components (eigenvalues > 1) were used in Canonical Variate Analysis.

3 Results

The results of ISSR analysis pointed out the genetic relationship between the two landraces. As shown in Fig. 1, the PCA highlighted a clear separation of the populations sampled in Conca Casale and Capracotta. In particular, along the first two PCs the total variance accounted for 45.57% and 18.04%, respectively. Molecular degree of differentiation between the two groups of populations (AMOVA) showed a significant molecular discrimination (PhiPT = 0.438; p = 0.001). Moreover, it resulted that the genetic variability was greater within (56%) than among (44%) groups of landrace populations.

The comparison of total seed proteomic maps of the Conca Casale and Capracotta populations revealed a total of 193 differentially expressed proteins. The biochemical data set (193 proteins) was subjected to ANOVA, to identify biochemical markers useful to distinguish lentil populations from different provenances. It resulted that 25 proteins were significant to discriminate Capracotta from Conca Casale lentils. PCA was computed on a correlation matrix, using these 25 significant proteins; the first two PCs explained 53.79%

and 11.62% of total variance, respectively, and the scatter plot of these two PCs indicated a clear distinction between the two groups of lentils (Fig. 2). Differences between the two landraces were tested by canonical variate analysis (CVA) computed on the extracted PCs. They were significantly discriminated (Wilks' $\lambda = 0.028$; df=5; $p < 0.0001$) as shown by the test of cross-validation (100% of cases were correctly classified).

The eight morphological variables were subjected to ANOVA in relation to population provenance. The two groups of lentil populations from Capracotta and Conca Casale were significantly discriminated by six morphological traits: seed density, roundness, volume, major axis length, perimeter and minor axis length. These six variables were used to compute a PCA: respectively, PC1 and PC2 explained 80.45% and 18.30% of total variance, highlighting a clear separation between landraces (Fig. 3). Then, the PCs were used in CVA and results indicated significant differences between the two population groups (Wilks' $\lambda = 0.047$; df = 2; sig.< 0.0001). Moreover, the test of cross-validations showed a high significance of the CVA reporting that 100% of cases was correctly classified.

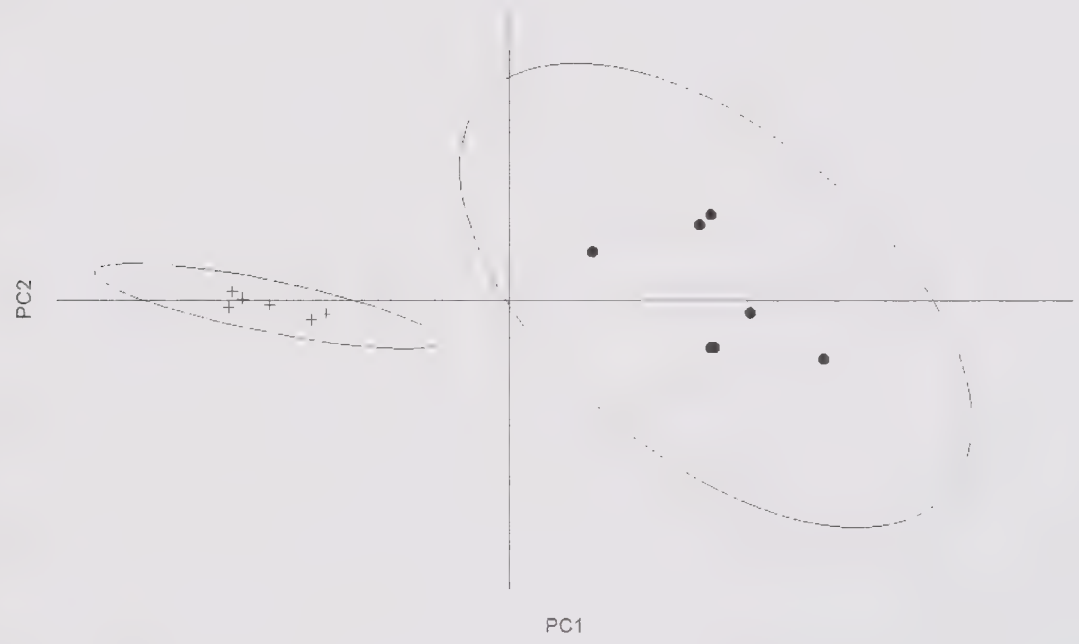


Fig. 1 – Scatter plot of specimens (cross = Conca Casale; point = Capracotta) ordered along the first two principal components; PCA from molecular data.

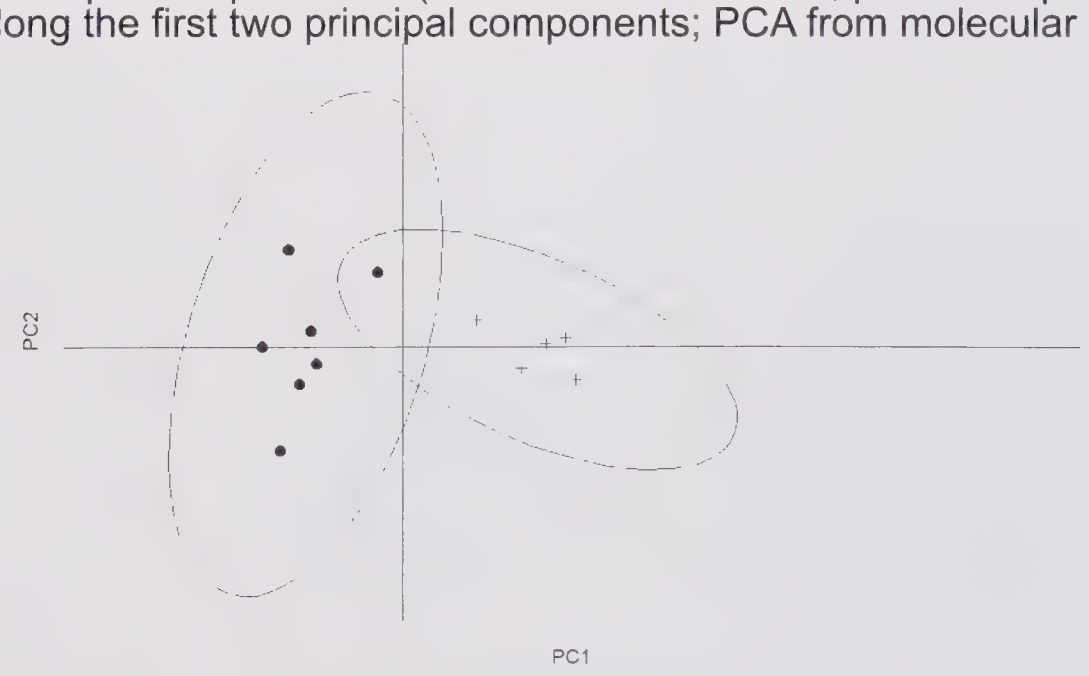


Fig. 2 – Scatter plot of specimens (cross = Conca Casale; point = Capracotta) ordered along the first two principal components; PCA computed on 25 proteins.

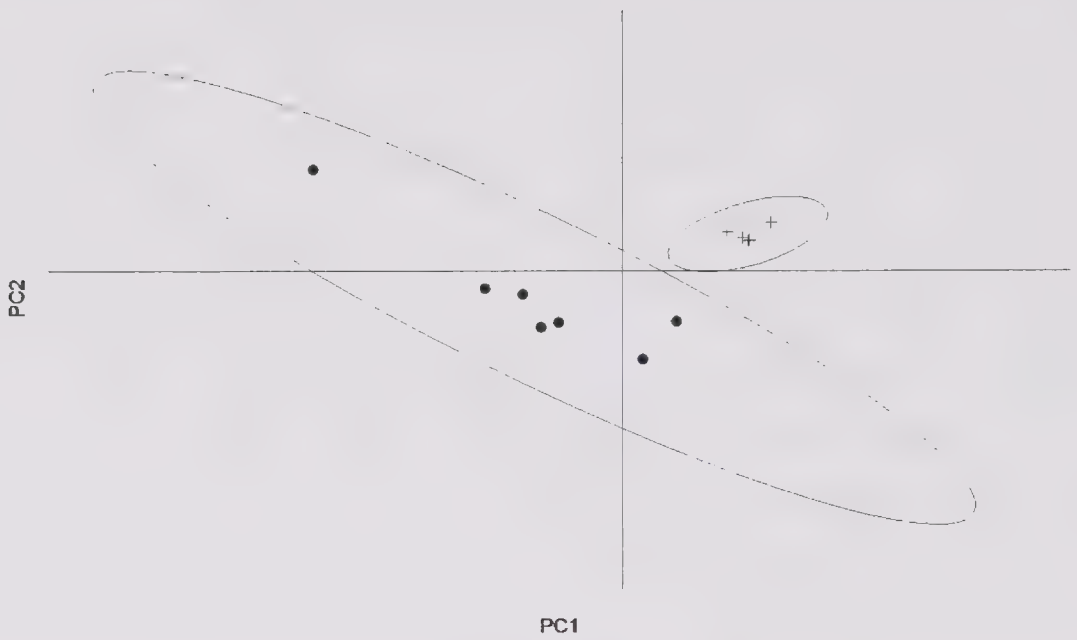


Fig. 3 – Scatter plot of specimens (cross = Conca Casale; point = Capracotta) ordered along the first two principal components; PCA computed on six morphological variables.

4 Conclusion

Autochthonous plant germplams, characterized by a wide genetic variability and high adaptation to different environmental conditions, are often more subjected to genetic erosion risks. In Italy, several different lentil landraces evolved thanks to the combination of different geographical characteristics.

The literature reports a wide variety of methods that have been used to investigate genetic similarities and relations among landraces of *L. culinaris* Medik.

Different methods have different powers of genetic resolution and provide different information: neutral DNA markers are useful tools to describe genetic relations in terms of time divergence [6], whereas phenotypic markers can provide information about adaptive responses to macro-environmental conditions [7].

In this study we used a combination of genetic and phenotypic analyses to characterize two autochthonous lentil landraces of two different provenances within a small region such as Molise.

The integration of genetic markers analysis with seed morphology and proteomic traits provided a high resolution approach to dissect lentil biodiversity [3]. The diversity between groups of populations, coming from two very close geographical areas, was well assessed and quantified. In addition, differences between the two local landraces were principally related to their sites of origin, where climate conditions and human activity may have selected the local accessions characterised by specific morphological and biochemical traits of seeds. Work is in progress to deepen the relation between these phenotypic markers and the environmental characteristics of the landrace provenance areas, and to identify the seed proteome markers.

References

[1] G. Ladizinsky, "The origin of lentil and its wild genepool", *Euphytica*, vol. 28, pp. 179-187, 1979.

[2] Y. Duran, R. Fratini, P. Garcia and M. Perez de la Vega, "An intersubspecific genetic map of

Lens.", *Theor. Appl. Genet.*, vol. 108, pp. 1265-1273, 2004.
[3] G. S. Scippa, D. Trupiano, M. Rocco, V. Viscosi, M. Di Michele, A. D'Andrea and D. Chiatante, " An integrated approach to the characterization of two autochthonous lentil (*Lens culinaris*) landraces of Molise (south-central Italy)", *Heredity*, vol. 101, pp. 136-144, 2008.
[4] G. Ladizinsky, "Wild lentils". *Crit. Rev. Plant Sci.*, vol. 12, pp. 169-184, 1993.
[5] A. R. Piergiovanni, "The evolution of lentil (*Lens culinaris* Medik.) cultivation in Italy and its effects on the survival of autochthonous populations", *Genet. Resour. Crop Evol.*, vol. 47, pp. 305-314, 2000.
[6] H. Thiellement, N. Bahrman and C. Damerval, "Proteomics for genetic and physiological studies in plants", *Electrophoresis*, vol. 20, pp. 2013-2026, 1999.
[7] J. L. David, M. Zivy, M.L. Cardin and P. Brabant, "Protein evolution in dynamical managed population of wheat: adaptative responses to macro-environmental conditions",*Theor. Appl. Genet.*, vol. 95, pp. 932-941, 1997.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 333-339.
ISBN 978-88-8303-295-0. EUT, 2010.

A model study for tardigrade identification

Roberto Bertolani, Lorena Rebecchi, Michele Cesari

Abstract — Using tardigrades from a single moss sample as a case study, we propose a new method for tardigrade species identification, which is often problematic, due to the low number of morphological characters. Identification at generic level was carried out on adults, while morphological analyses were performed on animals (LM) and eggs (LM and SEM), including hologenophores, vouchers used also for molecular analysis of COI mtDNA. This multi-approach method revealed the presence of three species of the "*Macrobiotus hufelandi* group" instead of the two species identified in a previous study. The validity of the method is shown, indicating that it could be applied to studies of problematic meiofauna taxa.

Index Terms — COI, DNA barcoding, morphology, Tardigrades, taxonomy.

1 Introduction

Tardigrades consist of more than 1,000 described species [1], [2] colonizing marine, limnic and terrestrial environments, including "hostile to life" and unpredictable habitats. In the seventies, a new evaluation of the intraspecific variability and new morphological characters for species identification were proposed [3], [4], [5], which led the number of tardigrade species to increase from less than 500 species described to that date to the current number. An example of this improvement in identifying species can be found in *Macrobiotus hufelandi*, the first described [6] and most commonly identified tardigrade species. What was considered a single species is currently represented by more than 25 species. Nonetheless, tardigrade identification at the species level is often problematic due to the low number of taxonomic characters. During our work it was not rare to find in the same moss sample more than one tardigrade species, not only in the same genus but also in the same species group, creating problems of species identification. For this reason we have begun to identify species by coupling a detailed evaluation of animal and egg shell morphology with DNA barcoding [7]. Using one moss sample as a case study, we propose a new method for tardigrade species identification and, in general, for identification of meiofaunal taxa whose morphological

The authors are with the Department of Biology, University of Modena and Reggio Emilia, 41125 Modena, Italy. E-mail: roberto.bertolani@unimore.it; lorena.rebecchi@unimore.it; michele.cesari@unimore.it.

characters are often very limited.

2 Material and methods

The moss sample was collected at Andalo (Central Alps; province of Trento, Italy, 46°N 10.133, 011°E 00.017, 1050 m) on a rock already examined by us (Fig. 1) [7], [8]. Several tardigrade species were present in the sample, belonging to different genera of eutardigrades (*Macrobiotus*, *Minibiotus*, *Ramazzottius*, *Milnesium*) and heterotardigrades (*Echiniscus*) but only the specimens belonging to the so-called "*Macrobiotus hufelandi* group" have been used in this study. Two species in this group, *Macrobiotus macrocalix* Bertolani & Rebecchi, 1983 (amphimictic) and *M.* cf. *terminalis* (parthenogenetic), were already found in the previous collections [7], [8].

Fig. 1 – The rock located in Andalo (Italy), with the moss patch that was used in the study.

Morphological analyses of paragenophore voucher specimens (*sensu* Pleijel *et al.* [9]) were carried out by mounting animals and eggs (that have species-specific shell processes) in Faure-Berlese fluid for light microscopy (LM) observations. Other eggs were fixed and dehydrated for scanning electron microscopy (SEM) analysis. Additional animals were stained with acetic lactic orcein for gender identification.

For molecular analysis, DNA was extracted from single entire animals. Some of these specimens were newborns hatched from isolated eggs, whose shells were mounted in Faure-Berlese fluid, obtaining hologenophore voucher specimens (*sensu* Pleijel *et al.* [9]). Amplification and sequencing of 684 bp of the COI mtDNA gene were carried out, following the procedures described in Cesari *et al.* [7]. Kimura 2-parameters distances between haplotypes were scored by using MEGA4 [10], while neighbor joining and maximum parsimony dendrograms were computed using PAUP* 4.01b10 [11], also using sequences retrieved from GenBank (EU244599; FJ435804-7; AY598773-5; FJ176203-17). For possible further investigations, a fragment of the moss sample was stored at -80°C.

3 Results

Observations of animals stained with acetic lactic orcein confirmed the presence of males and females morphologically attributable to *M. macrocalix* by the presence of a strong buccal armature, with thick crests and large bands of evident teeth and with a relatively wide buccal tube, also observed in mounted specimens (Fig. 2a). In addition, males were also found among the specimens characterized by a weaker buccal armature and narrower buccal tube (Fig. 2b).

Fig. 2 – Buccal-pharyngeal apparatuses (Faure-Berlese fluid, phase contrast). a: Strong buccal armature and wide buccal tube. b: Weak buccal armature and narrow buccal tube. Scale bars = 10 µm.

Molecular analysis of specimens and eggs belonging to the "*Macrobiotus hufelandi* group" revealed three clearly distinct haplogroups (a-c, Fig. 3), with very high genetic distances among them (Tab. 1).

		1	2	3	d
1	Haplogroup **a**				0.005
2	Haplogroup **b**	0.193			0.000
3	Haplogroup **c**	0.169	0.181		0.001

Tab. 1 – Kimura 2-parameters distances computed among (under the diagonal) and inside (column d) haplogroups.

A detailed analysis of egg shell morphology both of the hologenophores and of other voucher specimens (paragenophores) showed the presence of three types of eggs, all bearing processes as inverted goblets on the shell (Fig. 4). One type of egg exhibited high (9.6-10.7 µm) and wide (7.9-8.6 µm) processes with large (7.0-8.0 µm) smooth distal discs and very large pits located only around the process bases (typical of *M. macrocalix*) (Fig. 4a, d). The second type of egg was characterized by clearly smaller processes (7.4-8.4 µm) than

those of *M. macrocalix* and having an irregularly edged distal disc (6.3-7.0 µm in diameter) and a non-uniform reticulated egg shell with a thick meshwork (Fig. 4b, e). The third type of egg had small processes (5.0-5.3 µm in height) with a slightly irregular edge on the distal disc (4.7-5.2 µm in diameter), and a very uniform reticulated egg shell with a very thin meshwork (Fig. 4c, f).

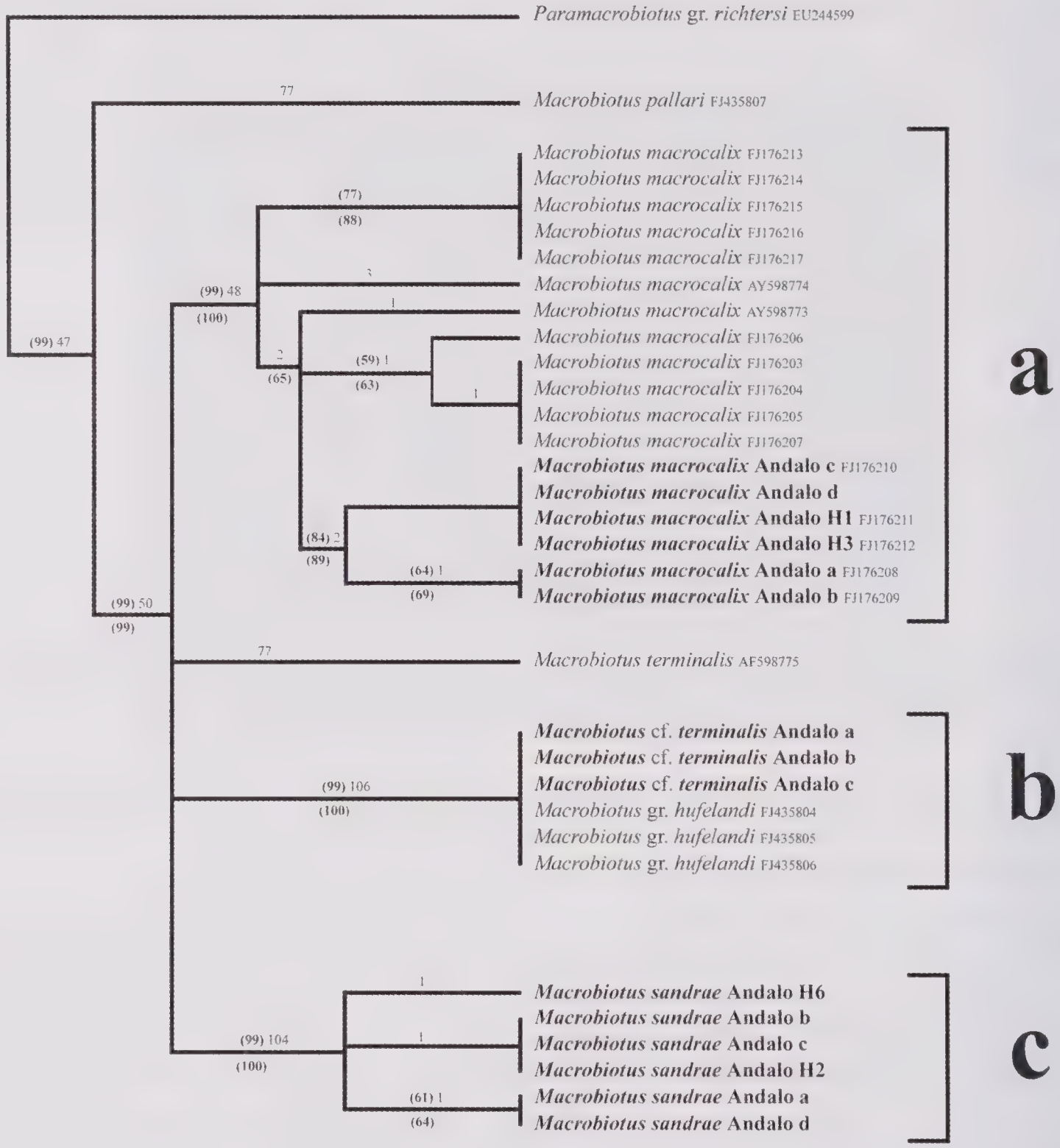


Fig. 3 – Dendrogram combining neighbor joining (NJ, ME score: 0.731) and maximum parsimony (MP, consistency index: 0.743; retention index: 0.920; rescaled consistency index: 0.684) analyses. Numbers above branches indicate mutational steps, while numbers in parentheses show bootstrap values computed after 2000 replicates (above branches: MP; below branches: NJ). **a**-**c** denote different haplogroups, while H denotes individuals for which hologenophore voucher specimens are available. Names in bold indicate specimens pertaining to the studied moss.

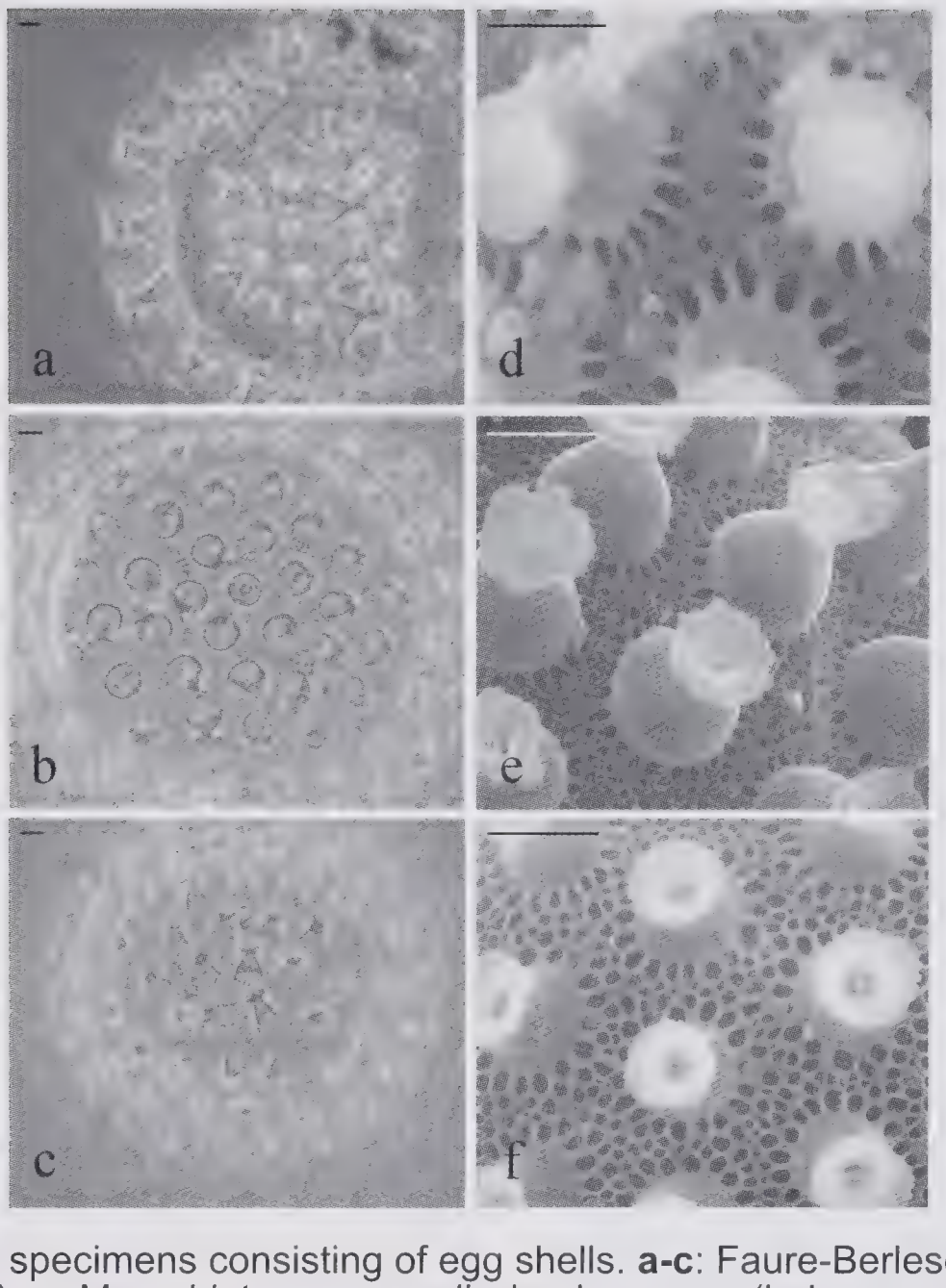


Fig. 4 – Voucher specimens consisting of egg shells. **a-c**: Faure-Berlese fluid (LM, phase contrast). a: *Macrobiotus macrocalix*, haplogroup a (hologenophore H1). b: *M.* cf. *terminalis*, haplogroup b (paragenophore). c: *M. sandrae*, haplogroup c (paragenophore). **d-f**: SEM (paragenophores). d: *M. macrocalix*. e: *M.* cf. *terminalis*. f: *M. sandrae*. Scale bars = 5 µm.

4 Discussion

The sex ratio analysis of the tardigrades belonging to the "*Macrobiotus hufelandi* group" in the moss sample revealed a much more complicated situation than that known from the literature [7], [8]. Nevertheless, by comparing the results of a detailed morphological analysis with those obtained by DNA barcoding, and in particular by sequencing the newborns' DNA and linking their sequences to the related egg shell shapes (hologenophores), the problem can finally be solved.

The distance values among the three different haplogroups are very high, far exceeding the 3% threshold and the 10x rule proposed by Hebert *et al.* [12], [13], [14], thus supporting the specific rank of the three haplogroups. Two species, *M. macrocalix* and *M.* cf. *terminalis* (currently being described as a new species), morphologically correspond to what was previously found on the same rock at Andalo [7], [8]. With regards to the third species, the animals look similar to the specimens of *M.* cf. *terminalis* (even through probably smaller), but the eggs are quite distinguishable and allow us to attribute them to *Macrobiotus sandrae* Bertolani & Rebecchi, 1983. This species is known to be amphimictic

[8], a situation consistent with the presence of males among the animals with a weaker buccal armature and narrower buccal tube.

5 Conclusions

The methods described here allow us to solve intricate tardigrade identification problems, validating our new approach based on linking morphological and molecular data. The use of voucher specimens, and in particular of the hologenophores, is critical for obtaining a correct species diagnosis. A hologenophore can also be obtained by culturing an isolated female until oviposition, mounting it as voucher and using its developing eggs either for molecular analysis and/or as further vouchers. Further information important for identification can also be obtained from other tardigrades, which can be photographed *in vivo* up to maximum magnification (100x objective) before being used in molecular investigations.

In our opinion, our multi-approach method for tardigrade identification can be easily applied to other meiofaunal taxa, whose few morphological characters can generate problems in species identification.

Acknowledgement

The authors wish to thank Diane R. Nelson, East Tennessee State University, U.S.A, for her help in the English revision and for her suggestions. This work was supported by a grant from the Fondazione Cassa di Risparmio di Modena (Italy) and the University of Modena and Reggio Emilia (Italy): "MoDNA project (Morphology and DNA): DNA barcoding and phylogeny of tardigrades, basic research and applications".

References

[1] R. Guidetti and R. Bertolani, "Tardigrade taxonomy: an updated check list of the taxa and a list of characters for their identification", *Zootaxa*, vol. 845, pp. 1-46, 2005.

[2] D. R. Nelson, R. Guidetti and L. Rebecchi, "Tardigrada". In: J. H. Thorp. A. P. Covich (eds.), *Ecology and Classification of North American Freshwater Invertebrates*, Elsevier Inc., Amsterdam, The Netherlands, pp. 455-484, 2010.

[3] G. Pilato, "Structure, intraspecific variability and systematic value of the buccal armature of eutardigrades", *Z. f. Zool. Systematik u. Evolutionforschung*, vol. 10, pp. 65-78, 1972.

[4] G. Pilato, "Redescription of *Haplomacrobiotus hermosillensis* May, 1948, and consideration of the genus *Haplomacrobiotus* (Eutardigrada)", *Z. f. Zool. Systematik u. Evolutionforschung*, vol. 11, pp. 283-286, 1973.

[5] G. Pilato, "On the taxonomic criteria of the Eutardigrada", *Mem. Ist. Ital. Idrobiol.,* vol. 32, Suppl., pp. 277-303, 1975.

[6] C. A. S. Schultze, Macrobiotus Hufelandii *animal e crustaceorum classe novum, reviviscendi post diuturnam asphyxiam et ariditatem potens*, C. Curths, Berlin, 6 pp. 1 tab., 1834.

[7] M. Cesari, R. Bertolani, L. Rebecchi and R. Guidetti, "DNA barcoding in Tardigrada: the first case study on *Macrobiotus macrocalix* Bertolani & Rebecchi 1993 (Eutardigrada, Macrobiotidae)", *Mol. Ecol. Resour.*, vol. 9, pp. 699-706, 2009.

[8] R. Bertolani and L. Rebecchi, "A revision of the *Macrobiotus hufelandi* group (Tardigrada, Macrobiotidae), with some observations on the taxonomic characters of eutardigrades", *Zool. Scripta*, vol. 22, pp.127-152, 1993.

[9] F. Pleijel, U. Jondelius, E. Norlinder, A. Nygren, B. Oxelman, C. Schander, P. Sundberg and M. Thollesson, "Phylogenies without roots? A plea for the use of vouchers in molecular phylogenetic studies", *Mol. Phylogenet. Evol.*, vol. 48, pp. 369-371, 2008.

[10] K. Tamura, J. Dudley, M. Nei and S. Kumar, "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA), software version 4.0", *Mol. Bio. Evol.,* vol. 24, pp. 1596-1599, 2007.
[11] D. L. Swofford, "PAUP* Phylogenetic analysis using parsimony (*and other methods)", Version 4.0b10 win32", Sinauer Associates, Sunderland, USA, 2002.
[12] P. D. N. Hebert, A. Cywinska, S. L. Ball and J. R. deWaard, "Biological identifications through DNA barcodes", *P. Roy. Soc. Lond. B Biol.*, vol. 270, pp. 313-321, 2003.
[13] P. D. N. Hebert, S. Ratnasingham and J. R. deWaard, "Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species", *P. Roy. Soc. Lond. B Bio.*, vol. 270, pp. 596-599, 2003.
[14] P. D. N. Hebert, M. Y. Stoeckle, T. S. Zemlak and C. M. Francis, "Identification of birds through DNA barcodes", *PLoS Biol.*, vol. 2, e312, 2004.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 341.
ISBN 978-88-8303-295-0. EUT, 2010.

DNA Barcoding of Philippine plants

Esperanza Maribel G. Agoo

Abstract — DNA barcoding is a technique that uses DNA sequence data for species-level identification, for analyzing phylogenies and interspecific variation and in population genetic studies. The DNA barcodes determined by the Consortium of the Barcode of Life (CBOL) to be the most effective in achieving these goals in plants are the plastid genes namely *trnH-psbA*, *rbcL*, *matK, accD, rpoB, rpoc1*, and *trn*L(UAA)-*trn*F(GAA). The goal of this barcoding project is to test the genes, *trnH-psbA*, *rbcL*, and *matK,* in identification and describing variation in some Philippine noteworthy indigenous plant groups such as orchids, gingers, aroids, cinnamons, and cycads. DNA are extracted and processed using standard protocol set by the CBOL. The DNA are then kept in a cold storage facility in the DLSU-CENSER laboratory. Voucher specimens are also collected and are now deposited in the DLSU-Manila Herbarium. Results of this study show that these candidate barcodes can successfully discriminate species including probable novelties of cycads, aroids, and cinnamons. Quantitative analysis also suggest that *rbcL* and *trnH-psbA* are very variable genes and can reflect greater interspecific variation thus are the more useful barcodes in these plant groups.

Index Terms — barcoding, Philippines, DNA, identification, flora.

The author is with the Biology Department, De La Salle University-Manila, 2401 Taft Avenue, Manila. E-mail: esperanza.agoo@dlsu.edu.ph.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 343.
ISBN 978-88-8303-295-0. EUT, 2010.

Molecular and ecophysiological characterisation of the Tunisian bee: *Apis mellifera intermissa*

Mohamed Chouchene, Naima Barbouche,
Lionel Garnery, Michel Baylac

Abstract — This study concerns the morphological identification, the molecular features and the eco-physiology of the Tunisian bee *Apis mellifera intermissa,* focused on 655 colonies from 7 populations: Kroumirie Moogod, North East Cap Bon, Ridge and Tell, high steppe, lower steppe, Atlas Chainon, Jeffara and Ouarra. The geometric morphometry of the interior wing of the bee shows polymorphism in size and shape. The size polymorphism is essentialy related to beekeeping practices. The characterization by means of a cytoplasmic molecular marker - mitochondrial DNA (DNA m t) - showed that the Tunisian bee originated from lineage **A**, which contradicts its membership to lineage **M** as demonstrated by a study based on biometric data only (Ruttner, 1988). There is a genetic polymorphism of the Tunisian bee in the presence of four haplotypes: A1, A8, A9 and A4. The distribution of the A4 and A9 haplotypes depends on ecological conditions. Foreign haplotypes are present in the region of Ghardimaou near the Algerian border (C7 haplotype). The study of some ecophysiological parameters in colonies of *Apis mellifera intermissa* from 5 sites showed that the Tunisian bee is endowed with a very marked disregard for all haplotypes (A1, A4, A8 and A9). However, we report the existence of a difference between these haplotypes in thermoregulation, oviposition and respiration of solitary bees. The temperature of the A1 and A8 haplotypes brood nest is around 36°C while the A9 and A4 haplotypes brood nest has a temperature of 34°C when weather conditions are extreme. The A4 and A9 haplotypes fall into hibernation, the temperature of the brood nest ranging between 22 and 28°C. The A1 and A8 haplotypes have a high tendency to lay A9 and A4 haplotypes, which however is variable, ranging from zero to average depending on climatic conditions. A study of respiration of isolated honeybees showed a difference in oxygen consumption between haplotypes A1/A8 and A4/A9 at low temperatures.

Index Terms — *Apis mellifera intermissa*, DNA, ecophysiology, Tunisia bee, haplotypes.

M. Chouchène, N. Barbouche are with the INAT, 43 Avenue Charles Nicolles, 1082 Tunis, Tunisia. E-mail: barbouche.naima@inat.agrinet.tn.
L. Garnery is with the Laboratoire Evolution, Génomes et Spéciation, UVSQ - CNRS. E-mail:garnery@pge.cnrs-gif.fr.
M. Baylac is with the MNHNP,36 Rue Geoffroy St Hilaire, 75005 Paris.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 345.
ISBN 978-88-8303-295-0. EUT, 2010.

Biological identifications through mitochondrial and nuclear molecular markers: the case of commercially important crabs from Indian EEZ

Sherine Sonia Cubelio, K. K. Bineesh, K. Raj, Suraj Tewari,
Achamveettil Gopalakrishnan,
Valaparambil Saidumohammad Basheer, Wazir Singh Lakra

Abstract — The ability of a 600- 700 base pair section of mitochondrial and nuclear molecular markers (COI, 16S & ITS-I) to provide species level identifications has been demonstrated for large taxonomic assemblages of animals such as insects, birds, fishes and crustaceans. In the Indian context, there had been no comprehensive attempt to determine the molecular sytematics, evolutionary relationships or phylogeny within the crabs of genera *Scylla, Portunus* and *Charybdis* which support the commercial fisheries. The present study is the first attempt to test the suitability of using a DNA barcode approach to discriminate accurately the edible marine crabs from Indian waters. Partial sequence of COI, 16S and ITS-I revealed distinct species specific profiles supporting the morphological data with low levels of intraspecific genetic diversity. Differentiation of two species of *Scylla*, *Scylla serrata* and *S. tranquebarica* using taxonomic tools is problematic especially based on the adult morphology. DNA barcoding using COI and other regions such as 16SrRNA and nuclear ITS fragment proved to be efficient in discriminating the species. The study revealed that mitochondrial and nuclear molecular genes will be an effective tool for discriminating species of commercial importance and thus aiding in scientific management of marine fishery resources.

Index Terms — molecular markers, crabs, fishery, Indian Ocean.

S. S. Cubelio is with the National Bureau of Fish Genetic Resources (NBFGR) Cochin Unit, CMFRI Campus, P.B. No. 1603, Kochi 682 018, Kerala, India. E-mail: nbfgrcochin@eth.net.
K. K. Binesh is with CMFRI, P.B. No.1603, Ernakulam, Kochi 682 018, Kerala, India.
The other authors are with NBFGR, Canal Ring Road, Dilkusha P.O., Lucknow 226 002, U. P., India.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 347.
ISBN 978-88-8303-295-0. EUT, 2010.

Barcoding Fauna Bavarica – Capturing Central European Animal Diversity

Lars Hendrich, Michael Balke, Gerhard Haszprunar, Axel Hausmann, Paul Hebert, Stefan Schmidt

Abstract — The Barcoding Fauna Bavarica (BFB) is an All Species Barcoding campaign ran by the Zoologische Staatssammlung in Munich and the Canadian Centre for DNA Barcoding (www.faunabavarica.de). Core funding comes from the Bavarian Ministry for Science, Research and the Arts and from Genome Canada through the Ontario Genomics Institute. The initial funding period is from 2009–2013. Bavaria has the highest biodiversity of all German states, with at least 35000 animal species reported, representing a significant portion of the central European species diversity. Ecoregions include high altitude biomes, foothill areas and forested lowlands. The Zoologische Staatssammlung (ZSM) is one of the largest German natural history research institutions. It holds the world's largest collection of Lepidoptera and Germany's largest Hymenoptera collection. Since mid-2009, the BFB project has contributed DNA barcode records from 7208 specimens representing 3000 species and is therefore, after less than one year, one of the most comprehensive sources for local DNA barcode data. The focus groups for the initial phase were Lepidoptera (1820 species barcoded), bees (316 species), ants (39 species) and aquatic insects (322 species). Work on these focal groups will continue during 2010, with the goal to complete 80% of the Bavarian focal group species by the end of the year. New focal groups are Diptera, Mollusca, all Vertebrata and terrestrial Coleoptera, targeting 2000 species in 2010. Most tissue samples come from specimens in the ZSM collection, and where this was not feasible from freshly collected and identified specimens. This rapid progress reflects the strong involvement of taxonomists throughout the process, which is one of our key missions. We have implemented a system which co-ordinates vouchers stored in our main collection, with tissues as well as DNA samples in our DNA bank.

L. Hendrich, M. Balke, G. Haszprunar, A. Hausmann and S. Schmidt are with the Zoologische Staatssammlung, Münchhausenstraße 21, 81247 München, Germany. E-mail: hendrich@zsm.mwn.de.
P. Hebert is with the Biodiversity Institute of Ontario, 579 Gordon Street, University of Guelph, Guelph, Ontario, Canada N1G 2W1, E-mail: phebert@uoguelph.ca.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 349
ISBN 978-88-8303-295-0. EUT, 2010.

Molecular techniques for identifying North Sea fauna

Thomas Knebelsberger, Sandra Ditzler, Silke Laakmann, Inga Mohrbeck, Michael J. Raupach

Abstract — Accelerated biodiversity assessment is the key to understanding the relationship between biodiversity and ecosystem functioning, especially in times of rapid climate change and habitat destruction. For the marine fauna of the North Sea, morphological species identification is impaired by the small size of many taxa, morphological convergence, intraspecific variation and larval stages which often elude morphological identification. Accordingly, the use of molecular methods presents highly promising tools for fast and accurate species identification. The aim of the new established research group "molecular taxonomy of marine organisms" at the German Centre of Marine Biodiversity Research is to test and develop molecular methods for the identification of the marine fauna of the North Sea, aiding efforts to monitor biodiversity patterns and changes. The research will focus on the analysis and identification of specimens using DNA barcodes, and environmental samples, in particular zooplankton, using next-generation DNA sequencing techniques. In addition it is planned to develop molecular methods for a fast and routine identification of larvae of selected invertebrate and vertebrate taxa of economic value.

Index Terms — molecular methods, DNA barcodes, North Sea, fauna.

The authors are with the Senckenberg am Meer, Deutsches Zentrum für Marine Biodiversitätsforschung, Südstrand 44, 26382 Wilhelmshaven, Germany TK. E-mail:tknebelsberger@senckenberg.de.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 351.
ISBN 978-88-8303-295-0. EUT, 2010.

DNA Bank Network – connecting biological collections and sequence databases by long-term DNA storage with online accession

Matthias Geiger, Nicolas Straube

Abstract — In times of increasing numbers of methods and tools for the molecular identification of organisms, it is inevitable that researchers have to deal with an additional flood of samples: the extracted DNA of organisms which are in the researcher's focus. Today, voucher specimens – the specimens from which DNA was extracted – have to be placed in adequate biological collections and organisms' sequence data can officially be deposited in online databases such as Genbank, often a prerequisite for publication of results in peer-reviewed journals. DNA extracts do not underlie such rules, but adequate housing of DNA extracts, especially of rare or difficult to obtain species, will be a major task in the near future.
The DNA Bank Network bridges the gap between natural history collections and molecular sequence databases by providing online references to analysed specimens and inferred molecular data. DNA samples are linked to their respective vouchers and inferred molecular data are stored in public sequence databases, facilitating taxonomic verification of molecularly analysed organisms.
We provide the opportunity for long-term storage of DNA in the DNA Bank Network, giving other reseachers the opportunity to access DNA for further projects dealing with the same organisms. In this way, multiple sampling can be avoided and there is a direct link between the three main sources of information, i.e. the sampled organism, the DNA, and the sequence data. Here we present the functioning and layout of the DNA Bank Network, which currently connects DNA banks of four research museums in Germany: the Bavarian State Collection of Zoology (ZSM), the Botanic Garden and Botanical Museum Berlin-Dahlem (BGBM), the German Collection of Microorganisms and Cell Cultures Braunschweig (DSMZ), and the Zoologisches Forschungsmuseum Alexander Koenig (ZMFK). Presently, the DNA Bank Network allows to access DNA samples of more than 35.000 DNA samples and 11.000 taxa.

Index Terms — DNA, databases, molecular identification.

The authors are with the Bavarian State Collection of Zoology, Münchausenstrasse 21, 81247 Munich, Germany. E-mail: dnabank@zsm.mwn.de.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 353-354.
ISBN 978-88-8303-295-0. EUT, 2010.

Mitochondrial DNA sequences for forensic identification of the endangered whale shark, *Rhincodon typus* (Smith, 1828): A Case study

Kavungal Abdulkhadar Sajeela, Chandran Rakhee, Janardanan Nair Rekha, Achamveettil Gopalakrishnan, Valaparambil Saidumohammad Basheer, Joe Kizhakkudan Shoba, Kizhakkudan Joe, Wazir Singh Lakra

Abstract — The whale shark (*Rhincodon typus)*, the largest fish in the ocean, has become susceptible to over-exploitation and has a global conservation status of 'vulnerable to extinction' as listed by World Conservation Union in the *Red list of threatened species*. The increase in demand for its meat, skin and fins in international trade is a severe threat to the animal and its indiscriminate capture will have to be taken seriously as they may have a major impact in the marine ecosystem. *Rhincodon typus* was nominated in Appendix II of Convention on International Trade in Endangered Species (CITES) in April 2000, to enable adequate regulation of trade of whale shark products. Whale shark (*Rhincodon typus)* is enlisted as one of the protected species in India and its fishing prohibited under Schedule I of the Indian Wildlife Protection Act, 1972, according to the Order No.1-2/2001 WL1 Dated 28.05.2001, Govt. of India, so as to conserve the species in Indian waters. still illegal fishing prevails in Indian waters and the catch is processed in the vessel itself and sold in markets as meat chunks. To curb the illegal trade and marketing of fishery products from whale shark, for devising good management practices and for the strict law enforcement, accurate and reliable species identification methods using molecular tools are of paramount importance. In an effort to establish a comprehensive identification data set, we have generated a species-specific partial sequence data of the mitochondrial genome of properly identified stranded whale shark samples, covering the 16S rRNA (546 bp),

K. A. Sajeela, C. Rakhee, A. Gopalakrishnan and V. S. Basheer are with the National Bureau of Fish Genetic Resources (NBFGR) Cochin Unit, CMFRI Campus, P.B. No.1603, Kochi 682 018, Kerala, India. E-mail: nbfgrcochin@eth.net.
J. N. Rekha, S. J. and J. Kizhakudan are with CMFRI, P.B. No.1603, Ernakulam, Kochi 682 018, Kerala, India.
W. S. Lakra is with NBFGR, Canal Ring Road, Dilkusha P.O., Lucknow 226 002, U. P., India.

Cyt b (541bp), COI (600bp) genes as the reference genetic profile helping in accurate identification of any body parts of the species. In the year 2008, flesh suspected as that of the Wildlife protected whale shark (*Rhincodon typus*) was seized from fishermen by the Forest Range Officer (Govt. of Kerala), Kannur, Kerala, India and was brought before the Judicial First Class Magistrate, Thalassery, Kannur, Kerala, India. The detailed sample analysis and confirmation of species was carried out at NBFGR Cochin Unit (R.P.330/08, dt 29. 09. 2008). Based on DNA sequencing of 16S rRNA(525bp) and COI (600bp) Cyt b(541bp) genes and comparing with the sequences earlier generated by NBFGR (FJ375724, FJ375725, FJ375726, FJ456921, FJ456922, and FJ456923), the suspected sample was identified as that of endangered Whale Shark (*Rhincodon typus*) and the result was communicated to the court. This is the first criminal case in India in which scientific evidence was sought in forensic identification of the meat of an aquatic organism enlisted in the Wildlife Protection Act of India and the DNA markers reiterated their ability to reliably identify product/meat sample of a species, thus helping in curtailing illegal trade of the endangered organisms.

Index Terms — DNA markers, whale shark, identification.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 355-360.
ISBN 978-88-8303-295-0. EUT, 2010.

An assignment-based e-learning course on the use of *KeyToNature* e-keys

Pencho Mihnev, Nadezhda Raycheva

Abstract — This article conceptualises a common approach to train teachers and university lecturers in integrating the use of e-keys produced by the EU-funded project *KeyToNature* in the design, development, and delivery of e-learning modules within their school practice. The approach is based on setting and administering learning scenarios developed around real practice assignments that have at their core the identification of organisms and end up with the creation of meaningful biodiversity information products, rooted in concrete environmental contexts. The learning products may even include the creation of specific sub-keys from the initially used e-keys. The paper briefly describes the application and the real testing of the approach in an experimental teacher training course conducted in Bulgaria within the *KeyToNature* project.

Index Terms — active learning, assignment, biodiversity, e-keys, e-learning course, identification tools, learning design, teacher training.

1 Introduction

The *KeyToNature* project shares a big variety of electronic identification tools and instruments for their handling in different modes and on different hardware and software platforms. Using and integrating these tools within different learning scenarios and contexts is a concrete task of the teachers. Some or all stages of the applied learning scenarios and the resulting final products are often not digitised, which has obvious negative effects. Another consideration related to the digitisation is the use of different e-tools not pertaining to a united system. This complicates the organisation of learning, leads to efficiency issues, and may lower the re-usage potential of the achieved results.

P. Mihnev is with BIKAM Ltd., 46, Oborishte str., Sofia 1505, Bulgaria. E-mail: pmihnev@gmail.com – N. Raycheva is with the Department for In-Service Teacher Training, Sofia University "St. Kl. Ohridski", 224, Tzar Boris III bld., Sofia, Bulgaria. E-mail: raycheva.n@gmail.com.

2 The Use of a Learning Management System as a Learning Platform for the Creation of Effective Scenarios

The identification process is "a means to an end", not "the end" itself. An implication about learning by using identification tools is that the identification needs to be integrated in wider contexts of meaningful learning scenarios.

Designing good scenarios requires good pedagogical skills, knowledge of the subject matter, adequate technological skills, and the availability of technological facilities for implementation. A scenario should also reflect the active learning and motivational findings researched and described by learning and motivation theories, e.g. [1], [2], and [3].

The identification process is *per se* an active learning. The keys themselves become more complex, integrating add-ons that facilitate re-usage and complex learning activities.

In order to maximise the positive effects of the e-keys, one can use the support of e-learning platforms. A newer and richer concept about the use of e-learning platforms is that **activities** and *not the content* are in the core of an e-learning platform.

This concept directly corresponds both to the active learning methods and to the creation of contexts which exploit positive motivation factors.

Another feature of all widespread e-learning platforms is the inter-linkage of all resources, activities, communication, organisation, and assessment tools within the platform. This enables more complex learning events and activities to be organised and conducted as a "learning whole".

The authors have designed an experimental curriculum for teacher training to develop and use e-learning modules for students having as their core activity identification processes based on the use of e-keys. The Moodle learning management system (e-learning platform) was chosen for the concrete course design.

The main teacher training target group consisted of biology teachers that have average ICT-skills, are able to work with e-keys, but are not trained to work with any e-learning platform.

We have chosen to use only a few, but very important and powerful e-tools of the platform (resources and activities) that can add real value in the implementation of meaningful learning scenarios for the use of e-keys.

3 The Teacher Training Curriculum – the Course Study Programme

The teacher training course curriculum consisted of two main parts: the first part puts the teacher in a student role to alter her/his viewpoint in experiencing learning. The second part trains teachers on how to construct and conduct an e-Learning course.

The first part is based on Assignment 1:

"As a trainee in the first part of this course you - preferably in a team of 2 to 3 persons - should:

- Produce an electronic "Profile of a tree", following a predefined structure, by using the information previously entered for that purpose by your team in the e-learning platform.
- Publish and present in an attractive and appealing way your electronic "Profile of a tree" by using the tools and instruments of the e-learning platform.
- Edit and complement the identification key with collected and/or personally developed material."

This assignment consists of activities in two different environments: in the field and in the classroom.

Activities in the filed include tree identification with the e-key and observation of the characteristics that are used to identify it. Activities in the classroom include working on the structure of a tree profile and editing a e-key. The tree profile describes the following information:

1. Name of the tree: Latin, Bulgarian; 2. Classification (levels of detail: on learner's own judgment); 3 Photos (minimum 3 taken by the learner, and 2 from the e-key) - they should present a natural view of tree, leaf - margin, upper and lower surface, flower and fruit; 4. Description (following a worksheet); 5. Importance for mankind; 6. Do you know that… (interesting facts/information about the species); 7. Additional information – personal comments (personal opinion); 8. References and resources used.

The second part is based on the Assignment: "Develop your own e-module on identification for your specific case (subject, grade, and students)".

The curriculum, structured in chapters and activities, can be reviewed in the folder "Materials" of [4].

4 The Course

An experimental course was conducted on July 8th and 9th 2010 at the Department of Information and In-Service Teacher Training of the Sofia University. The Department campus includes a small park in front of the buildings where the field work was carried out.

Twelve participants took part in the course – 8 Biology school teachers, 2 Biology trainers from a training centre, 1 university lecturer in Botany, 1 Science expert from a Regional Educational Inspectorate. The course was held by the authors of this article.

The entry level of the course participants with respect to their ICT knowledge and skills was as follows:

1. All participants were able to work with the e-keys of *KeyToNature*;
2. No participant worked previously with the Open Key Editor[1];
3. All participants had an "average" skill level for working with ICT, namely: Windows, e-mail, Internet, MS Word, MS PowerPoint;
4. Only one participant worked previously with an e-learning platform - Moodle

---

The Open Key Editor is a software developed within *KeyToNature*. It permits to edit already existing keys or sub-keys extracted from them, by changing the text, adding images, adding new species, changing the structure of the e-key, and even creating new e-keys from scratch.

- without using it after the training.

The aim of the course was to test the developed curriculum and to receive feedback on effectiveness and efficiency of the course.

The twelve participants were grouped into 5 permanent groups; the work was conducted by each group as a whole.

In broad terms, the course time frame was set in the following way:

1. First half of the first day – work in the field: trees' identification, gathering additional information (taking photos, observing the environment, taking notes), and filling-in the "terrain" part of Worksheet 1 (profile of the tree);

2. Second half of the first day – work as an user of the Moodle platform and of the e-learning course: collecting the requested additional information from Internet and from the e-key, entering the collected data in the interactive geographic map of Moodle and in the prepared course multimedia database; working with the Open Key Editor and developing a sub-key.

Homework: development of a short PowerPoint presentation about the profile of the identified tree.

3. First half of the second day – Design of e-learning modules in Moodle: e-Course setting, student enrollment, work with selected *resources* (labels, folders, hyperlinked text, access to different study materials); developing an e-course programme/syllabus.

4. Second half of the second day – work with selected *activities* in Moodle: setting up interactive geographic maps, developing a database, developing an assignment, setting up a glossary. Starting the design of each trainee's own e-learning module.

Homework: Full design and preparation of the e-learning module, ready for use by students.

It was very encouraging to see that all groups managed to perform well the required activities and to develop the products envisaged in the course.

The results of the work of each group from the course can be seen in the Bikam's *KeyToNature* Moodle platform [4]. Photos from the teacher training course can be seen at the web-address provided in [5].

5 Evaluation of the Teacher Training Course

5.1 COLLES - Constructivist On-Line Learning Environment Survey

We used COLLES [6] as one of the two survey instruments to evaluate the course. COLLES comprises 24 statements grouped into six scales. The six groups are Relevance, Reflection, Interactivity, Tutor Support, Peer Support, and Interpretation. The concrete survey questions grouped by categories can be reviewed at [7]. Graph charts with all survey results can be viewed at [4].

Important feedback from the teachers' answers was that:

1. In terms of the question categories the course scores higher than the middle value of occurrence in the scale ("Sometimes").

2. The highest, almost maximal, score in the survey is assigned to the course relevance category. According to J. Cole and H. Foster [8], p. 192, the Relevance

category is the most important with respect to the assessment of the course design.

5.2 Free-Form Opinion Survey

In the second, free-form opinion survey, participants were asked about the strong and weak sides of the course and its methodology.

The articulated strong sides were:

1. The power of the e-learning platform to offer electronic means that unite into a learning whole and serve the overall learning process.
2. The power of attractivity to students of the final products created within (or with the support of) the e-learning platform.

The most often mentioned weak sides were related to:

1. The eventual lack of sufficient hardware for the implementation of e-learning scenarios - laptops for field work, the availability of computer labs.
2. About the course delivery – the very short duration of the course – only 2 days. The participants didn't know that the course was intentionally compressed to last only two days in order to test the possibility of achieving the main goals in such a short training time.

6 Summary and Conclusions

The e-course proved to be a success, especially having in mind the intentionally imposed extreme constraints (unskilled trainees to work with e-platform, and short duration of the course).

The use of an activity-based e-learning platform as Moodle maximises the learners' motivation, effectiveness, and efficiency of both students' learning and teacher preparation. The teachers have to be trained accordingly, preferably initially through a simple, not overwhelming training that utilises only some, yet powerful, learning activities in the e-learning platform.

Two other important implications are that our curriculum & learning design approach proved to be:

1. *"e-key independent"*. It can be used with *any interactive identification key*, irrespective of the organisms it identifies.
2. *"e-learning platform independent"*, providing that the corresponding platform has a set of functionalities/instruments that can realise appropriate learning activities.

Furthermore, the opportunity to mix/blend the course could be considered, to organise its delivery partly face-to-face, partly by distance.

The course was prepared in levels, in a modular way. This article has discussed only the first, initial level of the curriculum, which was also tested with teachers in the course. The next two course levels are meant to update and upgrade the knowledge and skills of the teachers, to fully employ the functionalities and instruments of the platform.

As a next step in learning design and delivery inquiries, it will be interesting to test the delivery of an entirely distance-online course, following the outlined curriculum.

Acknowledgement

This work has been supported by the *KeyToNature* project, ECP-2006-EDU-410019, in the frame of the *e*Content*plus* Programme.

References

[1] R. L. Hanna, "Active Learning = Remembering = Learning", *Las Positas College, Livermore, California, USA*. http://lpc1.clpccd.cc.ca.us/lpc/hanna/learning/activelearning.htm, July 2010.

[2] S. Hidi, "Interest and Its Contribution as a Mental Resource for Learning", *Review of Educational Research*, vol. 60, 4, pp. 549-571, Winter 1990.

[3] S. Reiss, *Who am I: The 16 basic desires that motivate our actions and define our personalities*, New York: Tarcher/Putnam, ISBN 1-58542-045-X, 288 p., 2000.

[4] K*eyToNature* Moodle e-learning platform. http://k2n.bikam.com/moodle-new, July 2010.

[5] Photo gallery "Training *KeyToNature*". http://picasaweb.google.bg/k2n.bulgaria/TrainingKeyToNature#, July 2010.

[6] P. C. Taylor and D. Maor, "The Constructivist On-Line Learning Environment Survey (COLLES)" *Curtin University of Technology, Pert, Australia*. http://surveylearning.moodle.com/colles/, July 2010.

[7] P. C. Taylor and D. Maor, "Example COLLES (Preferred and Actual)". *Curtin University of Technology, Pert, Western Australia*. http://surveylearning.moodle.com/mod/survey/view.php?id=3, July 2010.

[8] J. Cole and H. Foster "*Using Moodle: Teaching with the Popular Open Source Course Management System*", 2nd Edition. O'Reilly Media Inc., USA, 2008.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 361-365.
ISBN 978-88-8303-295-0. EUT, 2010.

User needs for interactive identification tools to organisms employed in the EU-Project *KeyToNature*

Astrid Tarkus, Emanuel Maxl, Christian Kittl

Abstract — The EU-funded project *KeyToNature* is developing and optimising interactive tools for identifying organisms, making them suitable for being usable in the field of formal education across Europe. To define the requirements of the target audience, research was conducted in 11 partner countries during an initial project phase. Teachers and lecturers from primary schools to university level were asked to express their views about selected existing identification tools in a qualitative survey. The target audience was asked about perception, strengths and optimisation options, output channels and pedagogical application fields. The results showed that the adaption of the tools to the range of local organisms and the native language of the audience represents a fundamental step.

Index Terms — education, EU project, identification keys, *KeyToNature*, user needs.

1 Introduction

The 3-year EU-funded project *KeyToNature* deals with interactive tools designed to identify organisms. These software and web-based tools are to be incorporated within educational structures with the objective of improving knowledge of biodiversity. *KeyToNature* aims to provide easy access to identification tools, to optimize their educational efficiency and ease of use. A further objective is to provide for the interconnection of these tools, so that multilingual access is possible and usage across Europe as a whole is enhanced.

A total of 14 project partners from 11 countries – Austria, Belgium, Bulgaria, Estonia, Germany, Italy, the Netherlands, Romania, Slovenia, Spain and the

A. Tarkus and E. Maxl are with the Market and Usability Research Department of evolaris next level Competence Centre, Graz, Austria; E-mail: astrid.tarkus@evolaris.net, emanuel.maxl@evolaris.net.
C. Kittl is CEO at the evolaris next level Competence Centre, Graz, Austria; E-mail: christian.kittl@evolaris.net.

United Kingdom – are participating in this project [1], [2].

2 Background Situation

The success achieved through the use of educational tools in the classroom is determined to a large extent by the competence of the teachers and the pedagogical concepts they employ [3], [4]. In their international review of education, for example, Kugeman & Fisher [5] established that the systematic involvement of teachers is crucial for the success of ICT tools – from setting up the learning processes and creating pedagogical concepts at the start, to reviewing and verifying the contents and results provided by the learners.

During the *KeyToNature* Project start phase, a state of the art presentation of the tools was undertaken [6]; a selection of the tools that were to be used for research purposes was put forward for review.

As many aspects of the existing tools and prototypes for pedagogical use have not previously been considered in the international context, we decided to conduct a user requirements analysis as a first step towards involving teachers in the development of the identification tools.

This paper presents the results of our user requirements survey in connection with the *KeyToNature* project. Our main objective was to identify specific details of how the tools are perceived. What do teachers think about the general concept of the tools and their use within the curriculum? Which didactic framework is suitable for the identification tools and how can they be implemented in lessons? Which medium (e.g. mobile phone, website) is perceived as most appropriate?

3 Research Design

In order to meet the target groups' needs, we decided to concentrate on a user-oriented design approach [7]. In particular, data on the pedagogical and educational requirements, as these relate to the identification of biodiversity were gathered by means of qualitative analysis in all partner countries.

The target audiences were lecturers who were recruited by the project partners in their respective countries. At each educational level - primary, secondary and university - focus groups were formed in all 11 participating countries in order to obtain input for all end-user segments [8].

Where it was not feasible to form focus groups, qualitative interviews – face-to-face, per telephone or email – were used as an alternative [9]. The tools were presented and afterwards discussed. Specific guidelines that covered the survey questions were employed. The collected data were subsequently analysed by the partner countries and summarised in a detailed report.

A total of 219 teaching staff participated in the survey that was conducted in the period October 2007 to February 2008. Of these, 152 were interviewed within focus groups, 33 were surveyed in face-to-face interviews and 11 were surveyed by means of email.

4 Tools Presented

The material under consideration consisted of a collection of existing tools: these were presented to the target audience, where possible, either in form of online prototypes or as concepts using PowerPoint slides - depending on their development status.

The identification tools (i.e. software-based identification keys) use different techniques to identify organisms. These include dichotomous keys which provide identification based on only two different possible selection options per stage, and multi-criteria keys which enable users to select several characters at once. Another option is to provide a free input field so that users can search for results on the basis of their own entered search criteria (e.g. taxon name).

The presented tools were:

1. Walking with woodlice (primary school level): a simple web-based key to woodlice in the UK with a colourful interface.
2. Key to Trees and Shrubs (primary and secondary school levels): a web-based tool that makes extensive use of pictures and graphics.
3. Earthworm Survey (secondary school level): a dichotomous key that uses PowerPoint slides.
4. Key to the Flora of Val Rosandra (university level): this employs the same software and UI as the “Key to Trees and Shrubs”, but includes substantially more species (c. 1000).
5. E-Flora iberica (university level): a web-based key to the plants of the Iberian Peninsula with free text search option and a powerful browsing mode.
6. Bumblebee (university level): a key to bumblebees in the form of an interactive flash application.

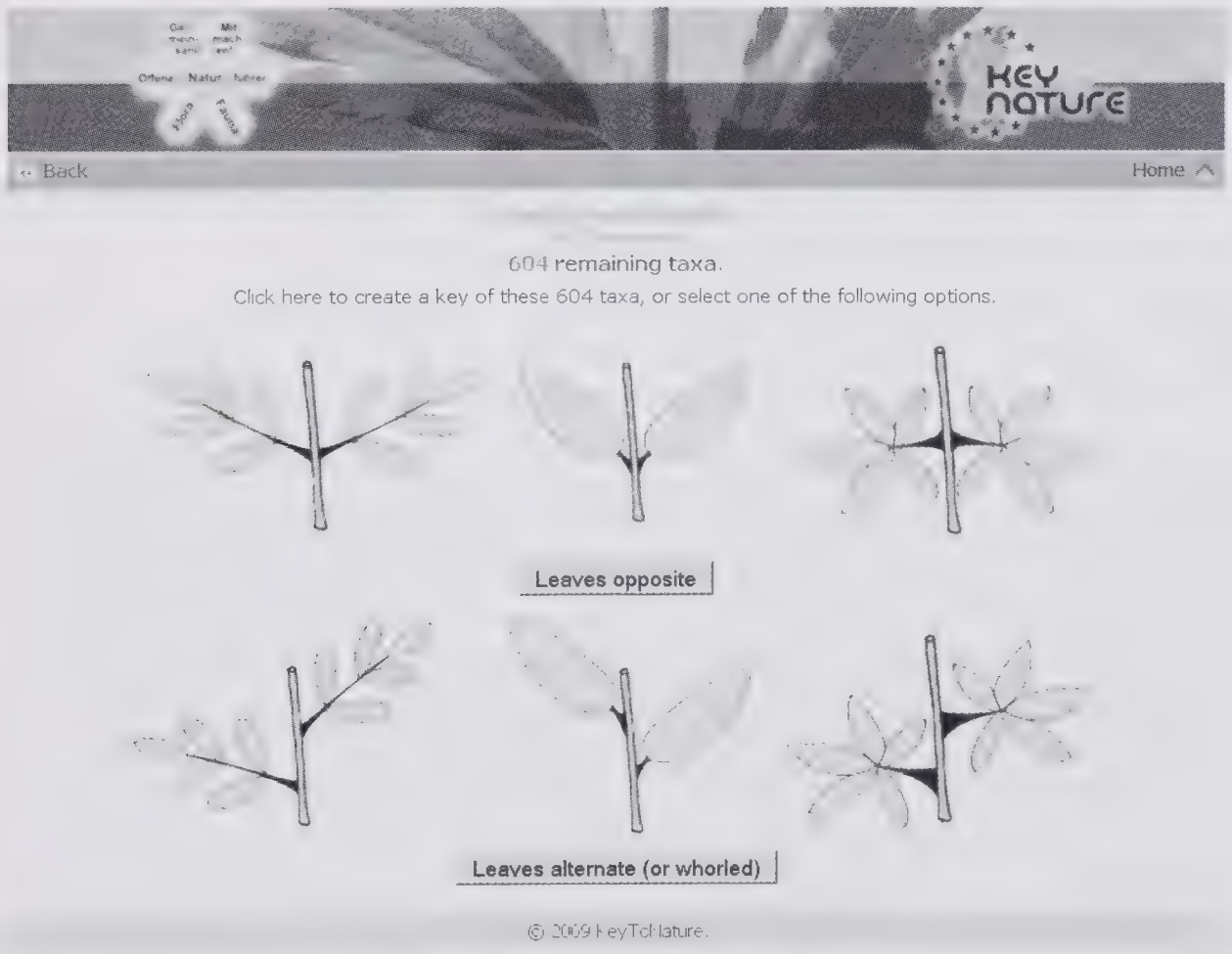

KEY nature
Back
Home
604 remaining taxa.
Click here to create a key of these 604 taxa, or select one of the following options.
Leaves opposite
Leaves alternate (or whorled)

Fig. 1 – Screenshot of a dichotomous key to woody plants (with options “leaves opposite” and “leaves not opposite”).
The current versions of the tools can be viewed at www.keytonature.eu.

5 Results

Perception of interactive identification tools: The dichotomous keys were perceived as an easily understandable, target group-oriented concept. Those surveyed considered that pupils/students would find this tool easier to use than the multi-criteria tools as they would be required to deal with less information at one time. The multi-criteria selection option should be provided for more advanced pupils/students, also because there was only limited pictorial information on individual characteristics available. The free text input option also met with a positive response, but was considered to be a more difficult tool to use for identification purposes, requiring more knowledge by the user with regard to the criteria that are crucial for differentiation.

Adaptation to the educational level: Our survey population considered that the woodlice key and the earthworm key would be suitable for a younger target group if the present application were improved - translations to native language and age-specific design. The idea of organism identification was generally seen as suitable at all levels, but the texts and designs needed to be tailored to the specific target audience.

Suitable media: In general, the fact that the target audience would need to be able to use a computer was not seen as problematic – even at primary school level – although these are not always available at all schools.

With regard to the media format, the CD-ROM was liked best by our survey population because this ensures there is no distraction by other web sites or services. Nevertheless, the web application was also perceived as positive in view of the better availability of updates, the platform it provides for online activities and its community-related aspects, such as the option for links with discussion forums, specialists etc. The mobile versions (for mobile phones or PDAs) were seen as outstanding in comparison with the other media because they can be used in the field; however, negative aspects were cited, such as the cost of data transfer and the limited screen size that makes it difficult to recognise organism details in images.

Pedagogical framework and educational applications: Several potential applications were identified; one that was frequently mentioned was the use for project work (at home, in the field or at school) at the primary and secondary school level, as current school curricula do not provide sufficient time to cover identification of organisms. Consequently, elective subjects would provide a perfect environment to work with identification tools. It was suggested that these tools could be used to present group projects in front of the class, thus helping improve pupils' presentation skills. However, one crucial aspect specified was that pupils at primary and secondary school level would need to be first instructed in the use of the tools by a teacher.

6 Conclusions

Several recommendations for optimisation based on the proposals made by our survey group were subsequently implemented. However, the survey itself had certain limitations. The tools were presented to the survey group only once,

and the individual teachers and lecturers were not given the option of trying out the tools over the long term in their educational institutes. Within the project progress, hundreds of tools were subsequently adapted to local requirements; local organisms were included and additional languages were integrated in the system with the help of some of the teachers and other associated members of *KeyToNature* who had been recruited for the project. Higher quality, high definition images and photographs and more interactive features were also added. At a subsequent phase of the project, an end-user evaluation was conducted by means of trials with students as subjects in order to investigate practicability, cognitive level of difficulty, and look-and-feel of the tools. A generally accessible platform was put in place in order to provide an arena for communication and for showcasing of ideas for the educational use of the tools. This has the potential to be further expanded and developed in future. Teachers and students were encouraged to participate by inputting descriptions, nomenclature and ecological data, distribution maps, images, sources and feedback. The purpose was to establish an international network of those interested in biodiversity and promote the dissemination of knowledge in this field.

Acknowledgements

The authors wish to thank all project partners that contributed in the user requirements analysis for valuable inputs and a great cooperation. *KeyToNature* is funded under the *e*Content*plus* programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable (ECP-2006-EDU-410019).

References

[1] Project website and platform for identification tools *KeyToNature*, http://www.keytonature.eu, 2010.

[2] S. Martellos and P. L. Nimis, "*KeyToNature*: Teaching and Learning Biodiversity: Dryades, the Italian Experience", *Proc. International Association for the Scientific Knowledge (IASK) Intern. Conf. Teaching and Learning*, pp. 863-868, Aveiro, 2008.

[3] D. H. Jonassen, *Modeling with technology: Mindtools for conceptual change,* Ohio, Prentice-Hall, 2006.

[4] J. M. Cox and C. Abbott, *ICT and attainment: A review of the research literature*, British Educational Communications and Technology Agency/Department for Education and Skills: Coventry and London, 2004.

[5] W. F. Kugemann and T. Fischer, *e-Learning 2005-2007: Eine Bestandsaufnahme der europäischen Beobachtungsplattform HELIOS* http://www.bibb.de/dokumente/pdf/a32_fernlernen_helios_fischer.pdf, presented at 5. BIBB-Fachkongress, Düsseldorf, 2007.

[6] P. L. Nimis, S. Martellos, G. Hagedorn, M. Brugman, M. Ferrer, A. Saag, T. Randlane, P. Schalk, B. Press, A. Barry and T. Trilar, *KeyToNature, Inventory of educational products* http://www.keytonature.eu/w/media/9/97/D.3.1_K2N_IDTools_Survey_Report.pdf, 2006.

[7] J. Nielsen, *Usability Engineering*. San Francisco, Morgan Kaufmann, 1994.

[8] T. L. Greenbaum, *Moderating Focus Groups. A Practical Guide for Group Facilitation,* London, SAGE Publications, 2000.

[9] G. Mey and K. Mruck, "Qualitative Interviews". In: G. Naderer and E. Balzer (eds.), *Qualitative Marktforschung in Theorie und Praxis Grundlagen, Methoden und Anwendungen*, Wiesbaden, Gabler, pp. 249-276, 2007.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 367-371.
ISBN 978-88-8303-295-0. EUT, 2010.

Teaching biodiversity with online identification tools from *KeyToNature*: a comparative study

Felicia Boar, Adelhaida Kerekes

Abstract — This paper reports on new pedagogical approaches in teaching biodiversity by using online interactive identification tools developed by the European project *KeyToNature*. A comparative educational study was carried out on two classes of students, revealing the educational value of the interactive identification tools. A systematic evaluation of both the educational processes and the acquired skills was conducted, and the results are presented here.

Index Terms — biodiversity, online identification keys, e-learning, assessment.

1 Introduction

Biodiversity represents the variety and variability of living organisms at the taxonomic and ecosystem levels [4]. Every species has a well-defined niche in natural ecosystems. Relations among species are complex and of different kinds, the disappearance of some species can bring imbalance in ecosystems [2].

Knowledge of biological diversity through curricular and extra-curricular activities is one of the main objectives of environmental education. Education on biodiversity is important for raising environment-aware citisens: the students have to be aware that every species has its own place and role in the maintenance of the ecological balance of the Earth, and that biodiversity safeguards the order of the planet affecting climate change, keeping the air clean, providing food, resources, medicines and potable water.

The teaching of biodiversity is a complex process: it needs a blend of classic educational methods (observation, simulation, explanation, field-work, visits to botanic gardens, zoos, natural history museums, etc.) with modern approaches such as the use of educational software and related questioning and team-

F. Boar is Professor of biology with the Vocational Environmental Protection High School, 400202 Cluj, Romania. E-mail: feliboar@yahoo.com.
A. Kerekes is Inspector of biology with the Cluj Inspectorate for Education, 400200 Cluj, Romanian E-mail: etti_aziza@yahoo.com.

research projects. These allow the forming of abilities in measurement, phenomenon recognition, ecosystem research, and – a basic task - identification of species.

2 Educational process

2.1 Didactic methods

Didactic methods – when used skillfully - lead to systematic and progressive enrichment of knowledge. They can be defined as "ways of acting, by means of which students can - independently or under the guidance of teachers - acquire knowledge, build up their skills, abilities and attitudes, develop their world and life outlook" [5]. Modern methods of teaching and learning with educational software (Fig. 1) are important in this context. They help students by providing a large amount of well-organized and structured opportunities for practicing skills in various fields of knowledge, allowing an easy and often playful simulation of processes and phenomena which are often hard or impossible to access directly [1], [3].

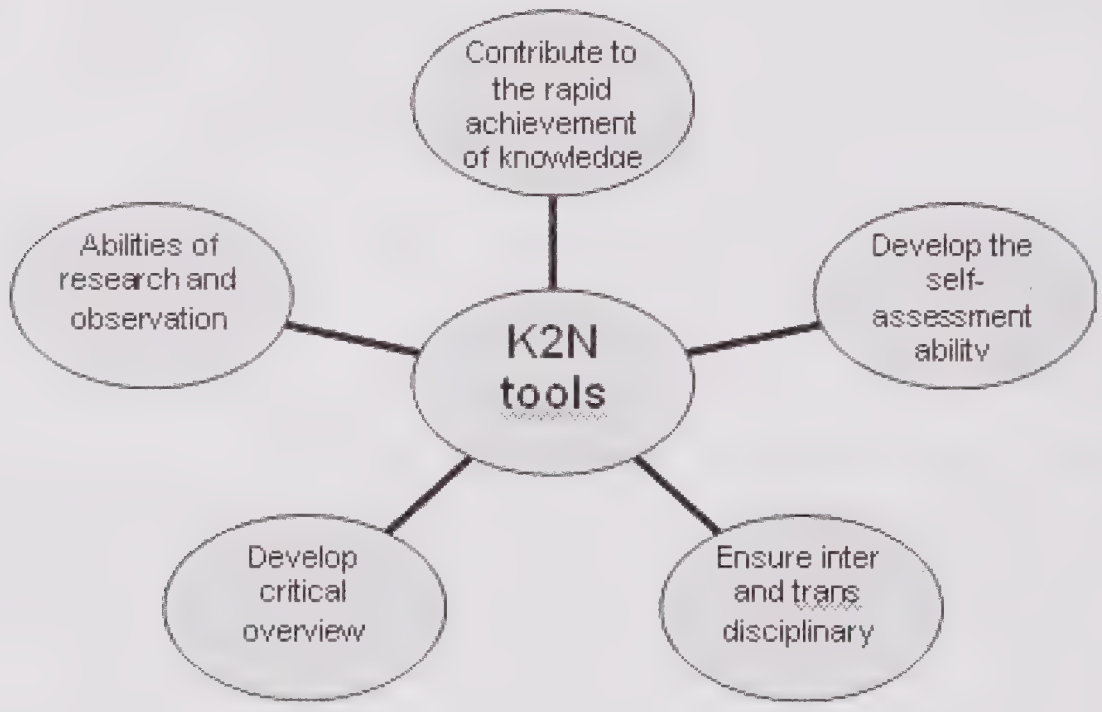

Contribute to the rapid achievement of knowledge
Abilities of research and observation
Develop the self-assessment ability
K2N tools
Develop critical overview
Ensure inter and trans disciplinary

Fig. 1 – Using the identification tools of *KeyToNature.*

We have conducted a comparative study in teaching the identification of Spermatophytes to two classes of students. Teaching methods were identical in the two classes (see Tab. 1): observation, modelling, a study tour, explanation, a botanical atlas, worksheets. For identification, however, a class used a pictorial botanical atlas, the other class used computer-assisted learning tools developed by the European project *KeyToNature* [6, 7].

2.2 Learning activities

The developed learning activities (see Tab. 1) aimed at teaching the general characteristics, recognising some representative species, discussing the ecology and economical/ecological importance of Spermatophytes.

Systematic group	**Topic / time**	**Training scenario**	
		Class A	**Class B**
Gymnosperms	General characteristics (1 hour)	• grouping students in four groups (collaborative tasks) • the students study the available biological resources • they observe the morphology of leaves and cones identifying the main characteristics of gymnosperms; these were written down afterwards on a dedicated worksheet	
	Identification & description of the species (1 hour)	• place: biology lab • usage guidelines • ***use of a botanic atlas to identify five species*** • use of worksheets for Class A	• place: computer lab • usage guidelines • ***use of a KeyToNature interactive key*** [7] • using the worksheets for Class B (e.g. Tab. 2)
Angiosperms	General characteristics (1 hour)	• grouping students in four groups (collaborative tasks) • the students study the available biological resources • they observe the morphological characteristics of leaves, flowers and fruits, identifying the main characteristics of Angiopsperms; these were written or drawn (leaves, flower) afterwards on the worksheet.	
	Identification & description of the species (1 hours)	• place: biology lab • usage guidelines • ***use of a botanic atlas to identify five species*** • worksheets	• teamwork - computer lab • usage guidelines • ***use of a* KeyToNature *interactive key*** • worksheets
Spermatophytes	field training (1 hour)	• visit in the park near the school • identification of the species studied in the lab • comparing the characteristics of different species (Gymnosperms vs Angiosperms) • discussing the adaptations to climate.	
Evalution (1 hour)		• evaluation test (see section 2.3 in this paper)	

Tab. 1 – Design of the teaching scenario.

The training activities were organised in a learning unit of six hours, combining: in-class training, visit outdoors, and evaluation. Tab. 2 shows a sample of the student worksheet for Class B (studying Gymnosperms).

No.	**Name of species**	**Bole**	**Leaves**	**Seeds/ Cones**	**Areal/ Importance**
1.	Taxus baccata	Straight ascending	Acicular, flattened	A single seed with a woody skin, covered by a red, fleshy Aryl	Hardwood used for fine carpentry and sculpture....
2.	Thuja orientalis	Straight	Small, scaly, have a dimple on the underside...	Several seeds included into a cone	Cultivated as ornamental plant in gardens and parks.
3.	...	...	...	...	...

Tab. 2 – Example from the worksheet used by students to fill in their responses.

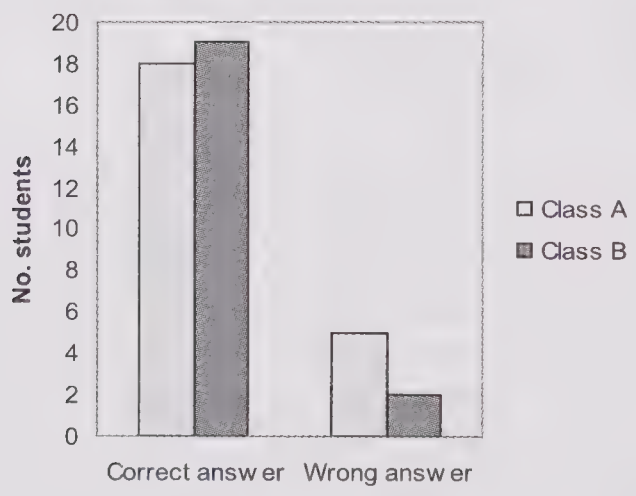
2.3 Methods and tools used for evaluation

Knowledge evaluation was carried out using different methods: systematic observation, self-assessment, correctness of identifications, as well as written paper. Both classes received the same test. The evaluation tool contained: short answer items, double choice items, pair items, questions on identifying and classifying, other structured questions. The evaluation was centered on: a) recognition of morphological characteristics of leaves and flowers of Gymnosperms and Angiosperms, b) description of the general characteristics of Spermatophytes, c) identification of species and their assignment to the correct systematic group.

Several questions concerned the recognition of morphological characteristics of leaves and flowers. The students in Class B (those who learned through the *KeyToNature* tools) identified better the morphological characteristics of leaves (Fig. 2) rathen than those of flowers (Fig. 3). This probablly depends on the fact that the interactive keys of *KeyToNature* were mostly based on leaf morphology.

The knowledge of the general characteristics of Spermatophytes was checked through short answers and double choice answers, with four statements on Gymnosperms and four on Angiosperms. For Gymnosperms, students in Class B had 25% more of correct answers compared with those of Class A, whereas for Angiosperms the results were similar in both claasses. This probably depends on the fact that Gymnosperms are a much smaller group with distinctive characteristics, which are easily appreciated by all students.

Finally, we have evaluated the ability of students to use morphological characteristics for identification and classification in the correct systematic group. The students had to identify four species of gymnosperms and four of angiosperms. Students from Class B had better results: most of them identified and classified correctly all species (Fig. 4). For example, 18 students correctly identified all species of Gymnosperms, 2 students correctly identified only 3 species, and 1 student only didn't identify any species (Fig. 4).



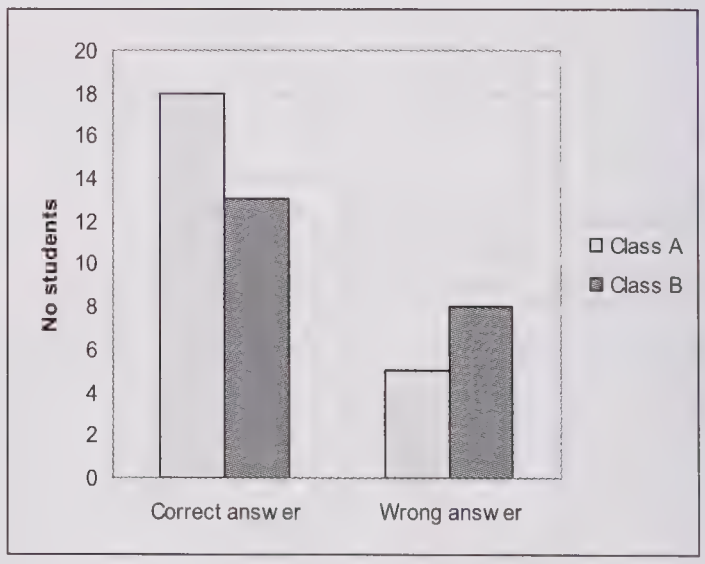


Fig. 2 (left) – Evaluation item: leaves morphology.
Fig. 3 (right) – Evaluation item: flower morphology.

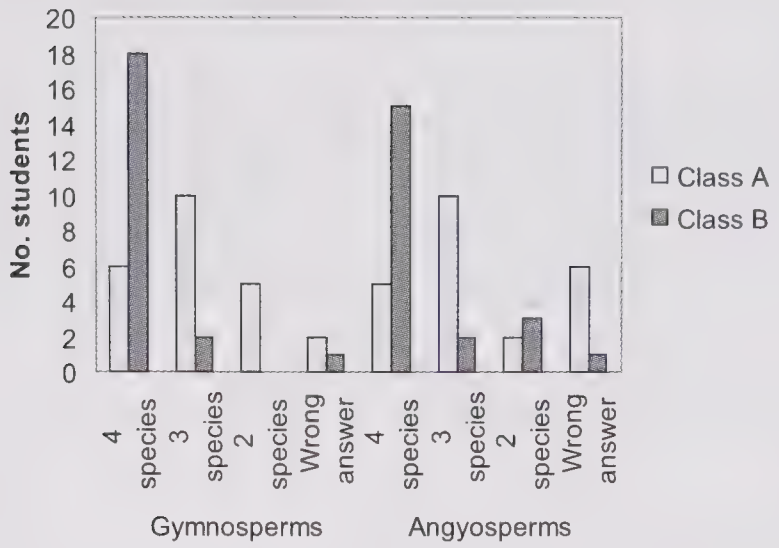


Fig. 4 – Evaluation item: systematic recognition and classification of species (number of correctly identified and classified species from a total of four).

These results indicate that the use of identification keys for the study of biodiversity allows students to develop observing, research and analytical skills, to enhance their intellectual work, and to improve their digital skills.

3 Conclusions

The study of biodiversity is one of the major goals of the scientific community and of worlwide policies. Educational goals can be achieved through trans-disciplinary approaches by bringing together traditional methods and innovative teaching. The use of computer assisted-learning such as the identification keys of *KeyToNature* proved to increase the quality and efficiency of the teaching–learning– assessment process.

Acknowledgement

This work was carried out during the testing activities in the framework of the *KeyToNature* Project (www.keytonature.eu), ECP-2006-EDU-410019. The authors wish to thank the project coordinator, Prof. Pier Luigi Nimis, for the development of the identification keys, and Prof. Mircea Giurgiu for providing access to the eLearning environment and for continuous local collaboration.

References

[1] L. Cohen and L. Manion, *Research Methods in Education*, 4th edition, Routledge, London and New York, 1994.
[2] V. Cristea and S. Denaeyer, *De la biodiversitate la OMG-uri?*, Eikon, Cluj-Napoca, 2004.
[3] L. Ezechil, *Prelegeri de didactică generală*, Paralela 45, Piteşti, 2003.
[4] V. Ghidra and M. Botu, *Biodiversitate şi Bioconservare*, Academic Press, Cluj-Napoca, 2004.
[5] M. Ionescu and I. Radu, *Didactica modernă*, Dacia, Cluj-Napoca, 2001.
[6] P. L. Nimis, S. Martellos, F. Boar, A. Kerekes, F. Crisan and M. Giurgiu, "A guide to the Pteridophytes of Romania", http://dbiodbs.univ.trieste.it/carso/chiavi_pub21?sc=412, July 2010.
[7] F. Boar, A. Kerekes, M Giurgiu and A. Homodi, "Spontaneous and Cultivated Gymnosperms", http://www.keytonature.eu/tools/gymnospermae/index.html, July 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 373-378.
ISBN 978-88-8303-295-0. EUT, 2010.

Digital Tools in the Botanical Garden of Madrid

Marina Ferrer, Esther García

Abstract — This paper describes the different roles of the Real Jardin Botanico of Madrid in the creation and use of digital tools related to biodiversity. The strongest point of this historical institution is the large amount of scientific data and items in its possession, which makes an excellent starting point for the creation of digital tools. The involvement of the Jardin Botanico in the creation of digital tools is centered on data providing, but some original tools have been developed as well, such as *E-Flora Iberica*, an ambitious system based on Flora Iberica. Another important role of the Botanical Garden of Madrid is that of testing the digital identification tools with users, since it receives about 500.000 visitors yearly, 50.000 of which are involved in formal education activities.

Index Terms — data provider, digital tools, dichotomous keys, education, Flora Iberica, wiki, workshops.

1 Introduction

The Botanical Garden of Madrid [1] has been involved in the use and creation of digital tools related to the identification of organisms during the last 3 years, as a partner of the European project *KeyToNature*. On one hand, it has created its own tools, on the other hand, it has served as a test-bed for identification tools developed in other contexts.

2 Data Providing

Being an institution with 250 years of history and a leading research center, the Botanical Garden of Madrid [1] owns an rich heritage of botanical resources: photographs, drawings, herbarium specimens, etc. Its most valuable contribution comes from *Flora Iberica* [2], a large project started in 1980, that gathers information on the flora of the Iberian Peninsula, with c. 4000 taxa described. The quality and quantity of the data produced in this project allow to create different tools for the identification of plant diversity. All of these data are freely

The authors are with the Scientific Culture Department, Botanical Garden of Madrid, Claudio Moyano 1, 28014 Madrid. E-mail: mferrer@rib.csic.es, estherr@rib.csic.es.

available online, and thus completely accessible to the broad public. Since the Iberian Peninsula is a biodiversity hotspot in Europe, this contribution covers about 60% of the European flora.

Dichotomous keys and taxon pages for 130 families and 732 genera, for a total of 3.560 species and 940 subspecies, were submitted to the online archive of *KeyToNature*, as well as 12.978 images of the flora of the Iberian Peninsula. In total, 242.330 metadata have been submitted.

3 Original identification tools

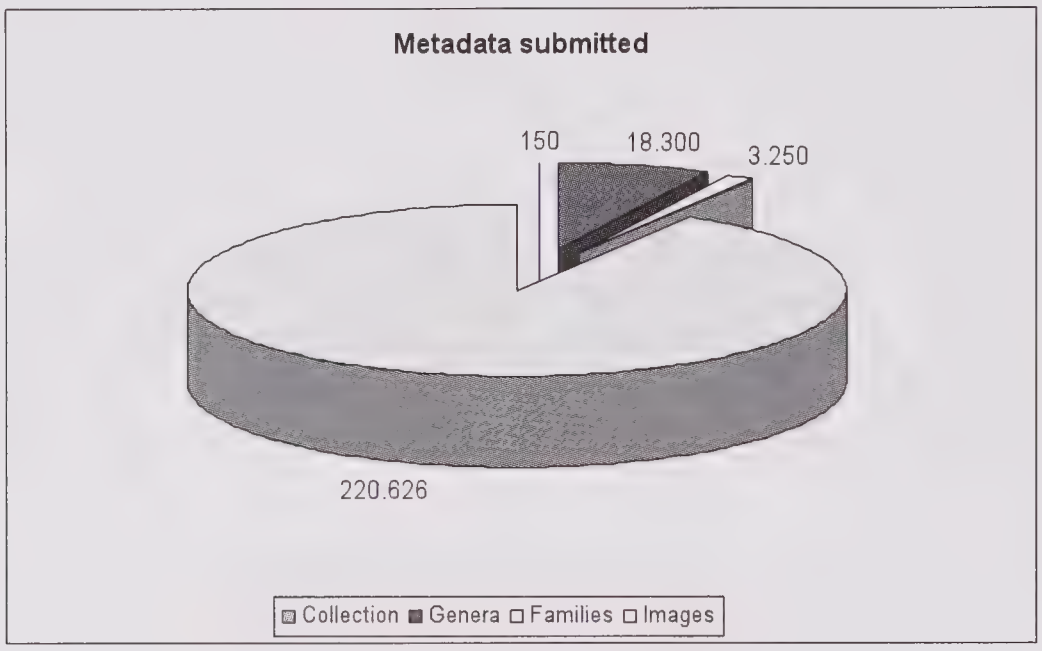


Fig. 1 – Distribution of metadata submitted to the *KeyToNature* online archive.

3.1 E-Flora Iberica

This is a digital identification tool which uses *Flora Iberica* as a raw material. Using *e-Flora Iberica* [3] one can obtain a long dichotomous key that includes all species, but can also create "mini identification keys" after setting up a series of filters such as province, family, etc. Another important feature is that the system can create a "minikey" out of some chosen species one wants to compare, providing a dichotomous key based on the differences among those species.

3.2. Wiki-keys

Using a wiki template developed in the framework of *KeyToNature*, two keys were created:

1. Wiki key of Gymnosperms of the Real Jardin Botanico - This is a dichotomous key of the gymnosperm species present in the Garden. It first leads us to genera (37 genera in the key), and for the genera that have more than one species (most of them), another subkey is shown, leading to the species. Information about the species and pictures (for most of them) are displayed in the taxon pages.

2. Wiki key for the ferns of the Flora of Equatorial Guinea [5] - Based on a modern flora of Equatorial Guinea, this wiki-key also follows a dichotomous structure ans is enriched with images.

2.3 The virtual assistant [6]

An experimental "virtual assistant" to identify plants was developed in the late phase of the *KeyToNature* project. This is a web application created from a very simple key: *Conifers of the Iberian Peninsula,* in which a "digital lady" literally "talks" to users, which can answer back by writing free text. Images can be implemented during the "conversation" to illustrate the process. The advantage with respect to a dichotomous key is that the user can skip several steps, depending on the amount of information she or he gives through the process of identification. The virtual assistant also has the quality of being amusing and attractive for youngsters.

4 Test Bed For Tools

The Botanical Garden of Madrid [1] offers excellent conditions for tools trial and testing, since it receives about 500.000 visitors yearly, 50.000 of them are involved in public education activities.

Several workshops centered on the use of digital identification tools were held at the Botanical Garden of Madrid [1] since the beginning of *KeyToNature*. Our main interest on these workshops was to develop learning experiences based on identification processes. The workshops participants' came from all grades of secondary school. 15 years old students were the most abundant age group (4th year Secondary School in Spain), with 449 pupils.

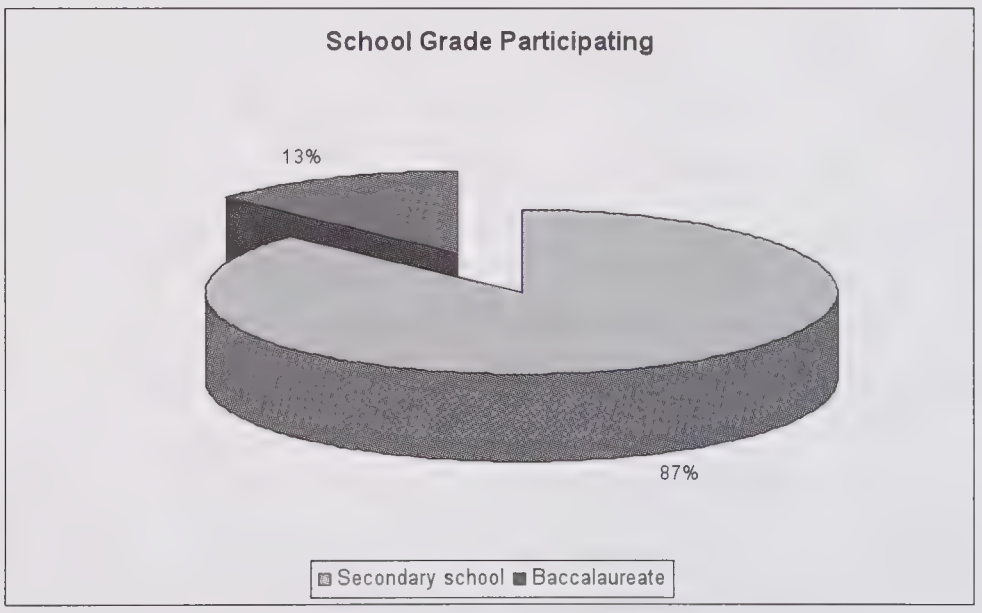


Fig. 2 – Distribution of the different school grades participating in the bioidentification workshops.

Most participants came from Madrid (65%), the rest from other municipalities (35%). 92% were from public schools, 7% from private schools and 1% from cultural centers.

4.1 Observation and Identification workshops

In total, 75 workshops took place under the framework of *KeyToNature* during 3 years. 69 of them were at the Botanical Garden of Madrid [1]. The workshops were carried with by groups of about 20 students accompanied by their teachers and assisted in the Garden by two previously trained instructors. The workshop consisted in 3 different stages: [7]

1. Brief explanation: The students were briefly introduced to the project and to the main features of the plants they would observe during the exercise: types of leaves, their disposal along the branch, whether they fall in winter or not, color and texture of tree bark.

2. Observation: Once the students knew what to pay attention at, they start drawing the plants or trees that the instructors had previously chosen. The identification tags of the plants were covered, to keep them "anonymous".

3. Identification: When sketches were finished, the group is conducted to the computer room of the Botanical Garden. There, they use an online identification key to identify the specimens. This key was produced by the *Dyades* project, the Italian branch of *KeyToNature*, for the Botanical Garden of Madrid [1], and in a simple and an advanced version [8]. The key is enriched with many images the pupils can compare with their drawings, in a didactic and entertaining process were the concepts learnt on the first two steps are efficiently fixed. This part lasts about 45 minutes.

4.2 Winter workshops

Thinking of the worst winter conditions in which the workshops could take place, a set of laminated leaves was prepared, which allowed students to have a close look at the leaves also in winter. During this exercise, even when they stood in front of a naked trunk, students had the opportunity to look at a well-preserved little branch. These waterproof resources served as a great support for the activity.

4.3 Other Workshops

Similar workshops took place elsewere: in the *Environmental Center Villaviciosa de Odón*, the *Botanical Garden of Alcalá* and in the Natural Area "*El Mesto*" near Madrid by the *Secondary Centre "San Agustín de Guadalix"*. These other centers had their own keys created especially for their needs, also in this case by the *Dryades* project [9].

Finally, the *Dryades* project also created another key for Spain: the *Key for Trees and Shrubs of Catalonia* [9], translated both in Spanish and Catalan, and posted in the web page of a public Catalan Centre of Science, the *Centre de Documentació i Experimentació en Ciències.*

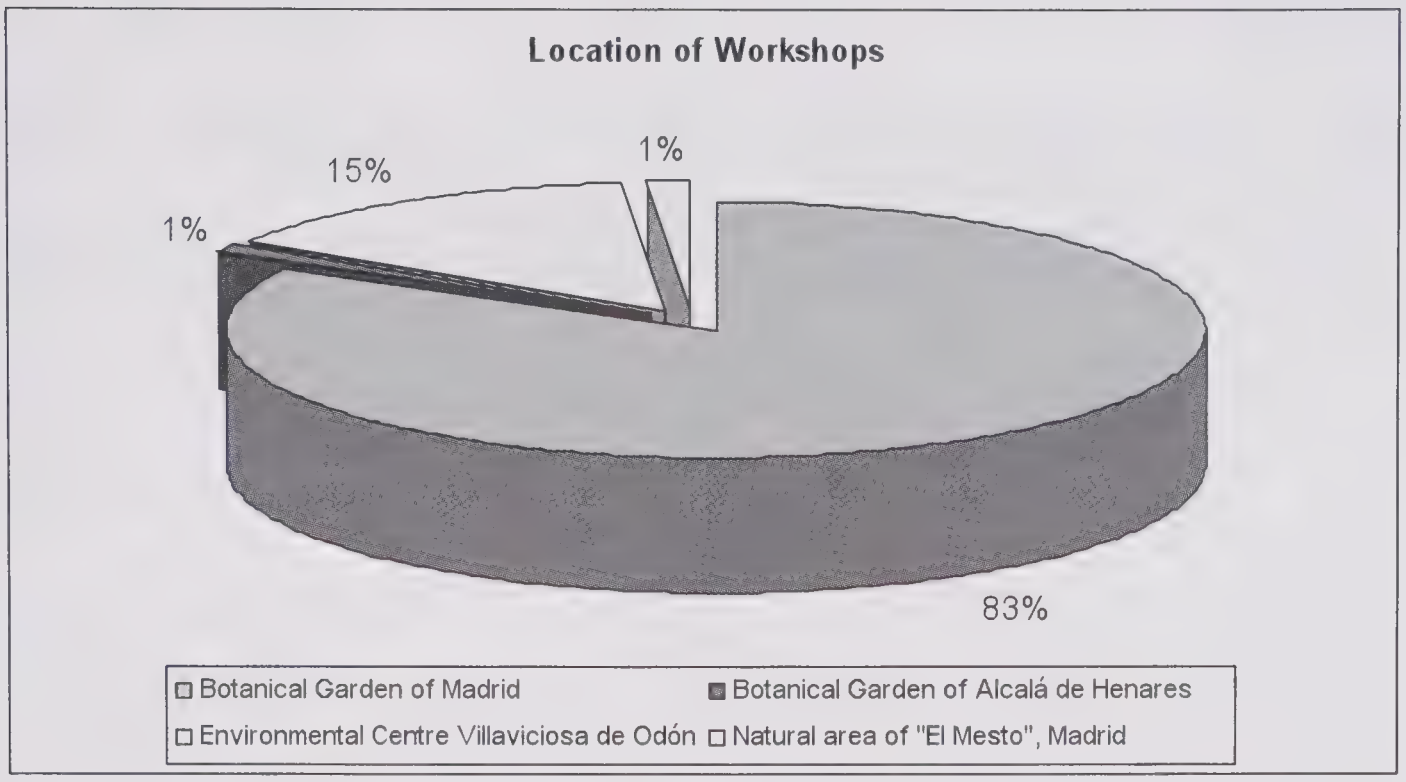


Fig.3 – Location of the workshops.

5 Keys on mobiles

We had the chance to try the *Dryades* key developed for the Botanical Garden of Madrid [1] on a PDA. This experience was carried out on the "Scientific Weekend", a science fair that took place in May 2010 at the *National Museum of Science and Technology*. An instructor was placed on a *KeyToNature* stand with two PDAs and a series of laminated leaves. The public that approached the stand (kids and adults) was able to try the key by using the mobile device. The activity attracted many visitors who found it very interesting.

6 Conclusion

The participation of the Real Jardin Botanico of Madrid [1] in the *KeyToNature* project as a data provider has put at disposal of the project the taxonomic, ecologic and biogeographical information for c. 4000 taxa of vascular plants from the Iberian Peninsula, as well as their images. This is an optimum data set for the development of digital tools for identification, due to the quality and rigour of the information generated by the Flora Iberica project along 30 years. Furthemore, the Garden has created the "e-flora" digital tool, and different identification keys in a wiki-format. It also performed as a tester of the digital tools generated under the *KeyToNature* project, with 73 experiences carried out with c. 1.400 students. The interest that these tools generate and their efficacy for the teaching of Natural Sciences was evident.

Acknowledgement

This paper was produced in the framework of the project *KeyToNature*, funded by the EU in the *e*Content*plus* programme.

References

[1] Botanical Garden of Madrid: http://www.rjb.csic.es/jardinbotanico/jardin/, 2010.

[2] Flora Iberica: http://www.floraiberica.org/, 2010.

[3] E-Flora Iberica: http://www.efloraiberica.es/eflora/, 2010.

[4] Wiki key for Gymnosperms of RJB: http://www.keytonature.eu/wiki/Clave_de_Gimnospermas_RJB, 2010.

[5] Wiki key for the ferns of the Flora of Equatorial Guinea: http://www.keytonature.eu/wiki/Clave_de_familias_de_Pteridophyta_de_la_Flora_de_Guinea_Ecuatorial_%28RJB%29, 2010.

[6] Virtual assistant: http://www.dialgraph.com/av/?idAsistente=1404151, 2010.

[7] Video "Observation and Identification Workshop" http://www.youtube.com/user/RJBCSIC#p/u/1/EDSVIJXsjEk, 2010.

[8] Links for the keys used at the Workshops: Simple version: http://dbiodbs.units.it/carso/chiavi_pub21?sc=119. Advanced version: http://dbiodbs.units.it/carso/chiavi_pub21?sc=165, 2010.

[9] Other Dryades tools created for Spain: Trees and shrubs of Catalonia: http://dbiodbs.units.it/carso/chiavi_pub21?sc=346. Key for the natural area "El Mesto" (Madrid): http://dbiodbs.units.it/carso/chiavi_pub21?sc=296. Key for the Nat. Area "El Forestal" (Vilaviciosa de Odón): http://dbiodbs.units.it/carso/chiavi_pub21?sc=311, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 379-381.
ISBN 978-88-8303-295-0. EUT, 2010.

Use of *KeyToNature* Identification Tools in the Schools of Slovenia

Irena Kodele Krašna

Abstract — This paper presents some results of the testing of several new interactive e-tools for learning and teaching biodiversity in the schools of Slovenia. The tools were developed in the frame of the ongoing *e*Content*plus* European project *KeyToNature*, in such a way to make them tailored to the needs of different educational users. We found out that identification of organisms with these tools is not only easy for students, but also for primary school children who had just learned to read.

Index Terms — biology, education, lesson, identification, *KeyToNature*, natural science.

1 Introduction

In the initial phase of the project *KeyToNature* we examined the need for identification tools in biology teaching. After this phase, we produced the first interactive identification keys in Slovene, and we tested them in elementary and secondary schools. In educational seminars for teachers we presented the new tools and the possibilities of their use. We showed teachers how our identification tools could be customized and adapted to their needs. Until now, over 50 tools were created in the Slovene language. All of them are freely accessible via the Internet and teachers are widely using them. Based on the responses of teachers, we greatly improved our identification tools, also developing several scenarios for their use in the educational system of the Country.

2 Methods

New interactive identification tools were tested in primary and secondary schools in Slovenia from June 2008 to June 2010. Teachers were invited to participate through the media and personal at meetings of study groups. Analysis of usability and applicability of our tools occurred on the basis of questionnaires sent back by teachers. Testings on the upper level of primary school and secondary schools were done by the teachers themselves, while on

The author is with the Slovenian Museum of Natural History, Prešernova 20, P.O.Box 290, 1001 Ljubljana, Slovenia.

the primary level (pupils aged 7–10) biologist (a *KeyToNature* project assistant in the Natural History Museum of Slovenia) helped the teachers. The usability and applicability of the identification tools were tested in different environments - in the computer lab and in the field using laptops and mobile phones. In the case of larger groups (over 40 students), for field-work we used printed versions of the identification keys. Pupils of the lower grades of primary schools have used smaller, customized keys (e.g. that to the plants of the lawn in front of the school in the village of Budanje).

3 Results

Completed questionnaires were returned by 19 teachers, 5 from the lower level and 12 from the upper level of elementary school, and 2 from high school teachers. A total of 710 students participated in the testing. The most used identification tool was an Interactive Guide to woody plants of Slovenia (31%). In 5 cases (26%) students worked with a key which was designed for a specific school (customized key). We mainly used the dichotomous interface of the keys (90%), because we soon realised that keys with a more complex approach - such as multi-entry keys - require more background knowledge and experience from students.

The interactive keys proved to be a very useful tool both in regular classes and in the other activities as science days and biology circles. After a brief introduction and explanation of basic terms used in the keys, pupils of the lower level of primary schools were capable to autonomously identify the organisms. While using a key, students learned about plants and animals specific to a certain habitat (meadow, forest, stream ...), observed their morphology, classified plants into the system. They also developed a functional way of reading, practicing cooperative learning and self-study. All teachers who participated in the survey stated that they would like to repeat the activity. Pupils enjoyed using the interactive keys because they had an active role and were independent at work. The activities proved to be successful independently from the form of the identification key which was used (interactive keys on the computer or printed on paper) and from where the activity took place (in the classroom or outdoors) (Fig. 1). The vast majority (80%) of teachers noticed that students were more interested in biodiversity after having used our keys, and several students even used them at home with their families.

Fig. 1 – Activities with the *KeyToNature* identification keys (a – pupils of the lower level of primary schools with a printed key, b – students of the higher level of primary schools with a stand-alone identification key on laptop, c – secondary school student working with an online identification key).

4 Conclusion

The identification tools – once adapted to users' needs – proved to be well suited for self-study or cooperative learning. They enable individualization and differentiation of lessons. The students like them very much, and the keys do not only add variety to lessons, but also make learning and the achievement of educational objectives easier: knowledge acquired in this way is more lasting.

Acknowledgement

I wish to thank Tomi Trilar for giving me the opportunity of working in *KeyToNature*. I am very grateful to The University of Trieste (Italy), especially Pier Luigi Nimis and Stefano Martelos, who have created for us several keys in the Slovene language. I would like to thank also Sonia Hetzner and Gerd Schmidt, who prepared the questionnaires, and the teachers who contributed in testings and gave us important feedback information. This paper was produced in the framework of the the project *KeyToNature* (www.keytonature.eu, ECP-2006-EDU-410019), funded in the e*Content*plus Programme.

References

[1] V. Trošt Vidic, "Learning about grassland plants in Podnanos, Slovenia", *KeyToNature* Teacher's Handbook, http://www.keytonature.eu/handbook/Learning_about_grassland_plants_in_Podnanos,_Slovenia, 2010.

[2] S. Mozetič, "Learning about grassland plants in Šempeter", *KeyToNature* Teacher's Handbook, http://www.keytonature.eu/handbook/Learning_about_grassland_plants_in_%C5%A0empeter, 2010.

[3] N. Šefer, "Trees and shrubs around the school", *KeyToNature* Teacher's Handbook, http://www.keytonature.eu/handbook/Trees_and_shrubs_around_the_school, 2010.

[4] M. Žohar, "Trees and shrubs in Laško", *KeyToNature* Teacher's Handbook, http://www.keytonature.eu/handbook/Trees_and_shrubs_in_La%C5%A1ko, 2010.

[5] A. Sodja, "Natural science lessons and identification keys", *KeyToNature* Teacher's Handbook, http://www.keytonature.eu/handbook/Natural_science_lessons_and_identification_keys, 2010.

[6] K. Prosen, "Let's learn about the families of vascular plants", *KeyToNature* Teacher's Handbook, http://www.keytonature.eu/handbook/Let%E2%80%99s_learn_about_the_families_of vascular_plants, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 383-387.
ISBN 978-88-8303-295-0. EUT, 2010.

New key-tools for pollen identification in research and education

Jade Dupont, Nathalie Combourieu Nebout, Jean-Pierre Cazet, Florian Causse, Régine Vignes Lebbe

Abstract — Pollen ID offers a free and easy access to various palynological information and compiles in the same web-space a pollen database and different services through a friendly user interface. Pollen ID proposes, or will propose, pollen and plant descriptions, terminology learning with an illustrated glossary and interactive images, identification keys, pollen analysis, pollen diagram construction, links with vegetation and climate data. The Pollen ID project is presently restricted to the European and Mediterranean geographical area, but it will be extended to other regions as well. This project is still in progress; its content and user interface – presently in French - will be soon available in English. In its final shape, the Pollen ID project will include palynological applications such as pollen determination tests, several original pollen analysis exercises with representations in diagrams and an easy interpretation of vegetation and climate. Pollen ID is accessible on http://lis-upmc.snv.jussieu.fr/pollen/.

Index Terms — pollen, free access key, identification.

1 Introduction

Palynology is commonly used in allergology, ecology, environmental reconstruction, climatology, and geology. Recently, it has been added in the current college program in France. Pollen identification by using books and online database is now largely used in palynology. Nevertheless numerous websites do not provide resource access for a large audience, from school education to research. Moreover, most of the websites do not link pollen data with the plant description and do not associate pollen applications to the descriptive content (our preliminary review has selected and analysed the

Jade Dupont was a master student of the University Paris VI. E-mail: j_r_xen@yahoo.fr.
Nathalie Combourieu Nebout, Jean-Pierre Cazet are with the LSCE - Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CNRS/CEA/UVSQ, Domaine du CNRS, avenue de la Terrasse, F-91198 Gif sur Yvette cedex, France. E-mail: Nathalie.Nebout@lsce.ipsl.fr.
Florian Causse and Régine Vignes Lebbe are with the LIS - Laboratoire Informatique et Systématique, UMR 7207 CNRS/MNHN/UPMC, MNHN Département Histoire de la Terre, CP48, 57 rue Cuvier, 75005 Paris, France.

content of 16 websites dedicated to the pollen description and identification). The Pollen ID project tries to take the challenge to fill this gap.

2 What is Pollen ID ?

Pollen ID offers a free and easy access to various palynological information and compiles in the same web-space a pollen database and different services through a friendly user interface. Pollen ID proposes, or will propose, pollen and plant descriptions, terminology learning with an illustrated glossary and interactive images, identification keys, pollen analysis, pollen diagram construction, links with vegetation and climate.

2.1 Pollen ID components

Resources are stored in two types of information systems: a relational database (MySQL) and knowledge bases for descriptive data (Xper^2) (Fig. 1).

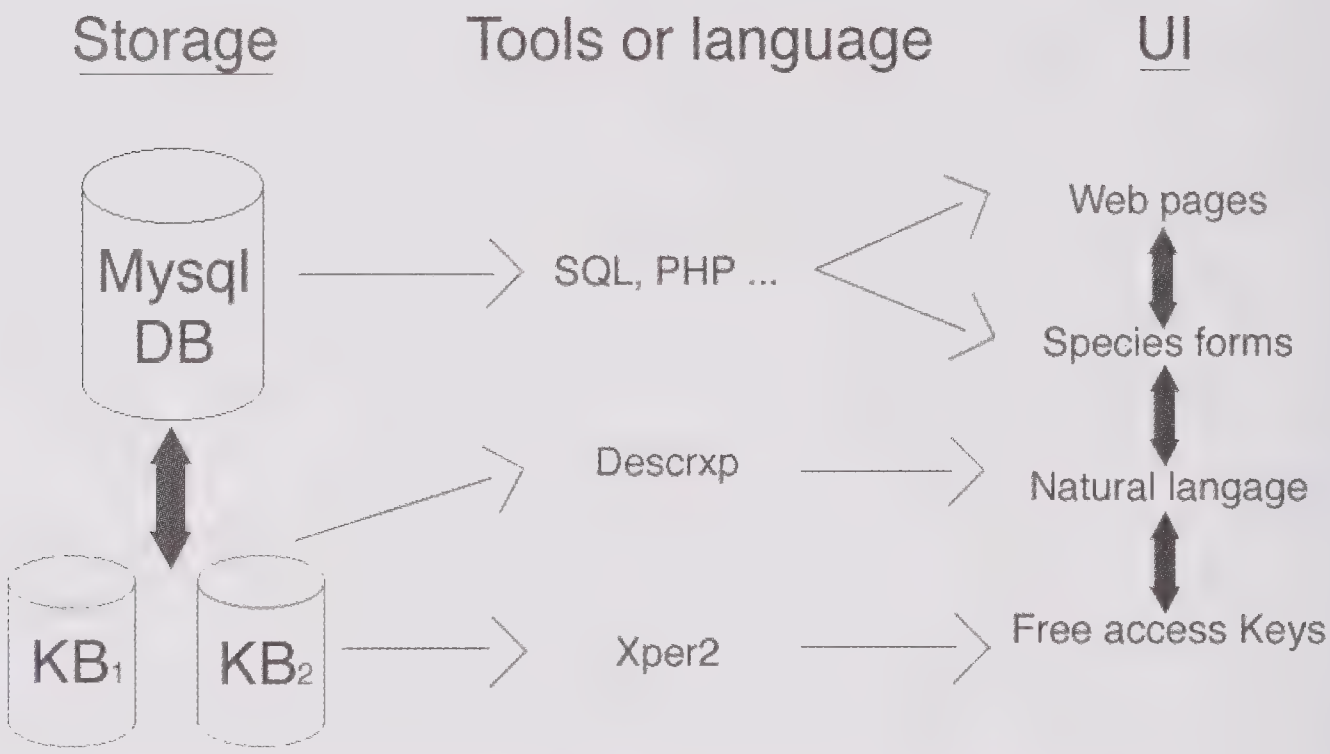


Fig. 1 – Pollen ID information technology architecture.

The relational database stores 1) nomenclatural data, plant information, photos and movies for the taxa, 2) definitions, pictures with image maps for the glossary terms, and 3) texts for the UI.

The descriptions of pollen are stored in two Xper^2 knowledge bases, and presently concern 94 Mediterranean species documented by texts, images and 67 videos:

- The first knowledge base is dedicated to beginners for training on pollen identification. 56 descriptors were selected, documented by text and images.
- The second one is more exhaustive and designed for expert users. It includes 115 descriptors with definitions and images.

The web pages are then dynamically constructed from the different resources (see Fig. 4).

3 Pollen ID use

The Pollen ID website user interface provides an original and large access to the complementary resources. Through the interactive buttons of the Home page, the user will discover numerous information from generalities on palynology, pollen descriptions and images to applications (in future developments) (Fig. 2)

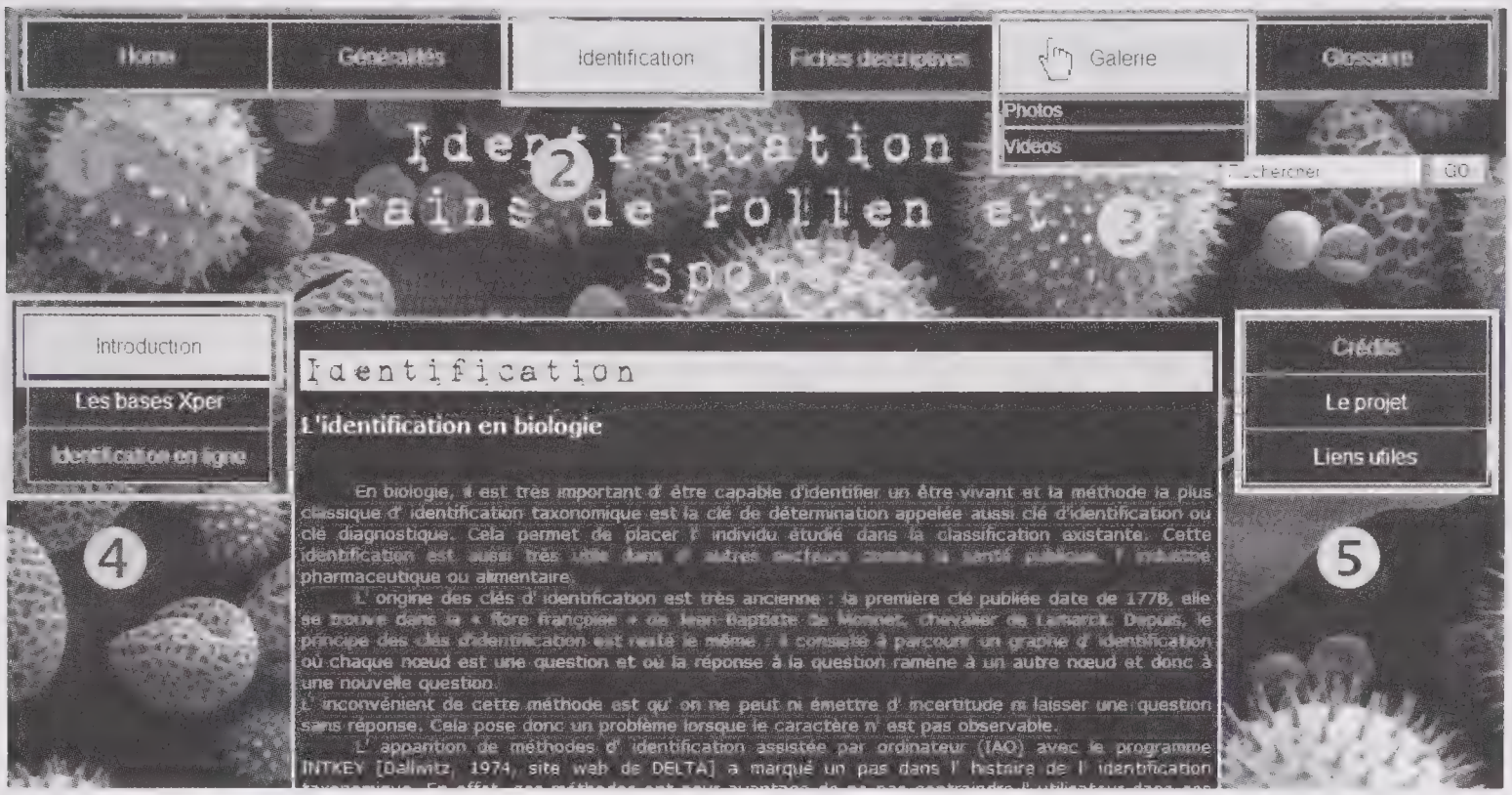

Home
Généralités
Identification
Fiches descriptives
Galerie
Glossaire
Photos
Videos
Chercher
GO
Introduction
Les bases Xper
Identification en ligne
Identification
L'identification en biologie
En biologie, il est très important d'être capable d'identifier un être vivant et la méthode la plus classique d'identification taxonomique est la clé de détermination appelée aussi clé d'identification ou clé diagnostique. Cela permet de placer l'individu étudié dans la classification existante. Cette identification est aussi très utile dans d'autres secteurs comme la santé publique, l'industrie pharmaceutique ou alimentaire.
L'origine des clés d'identification est très ancienne : la première clé publiée date de 1778, elle se trouve dans la « flore française » de Jean Baptiste de Monnet, chevalier de Lamarck. Depuis, le principe des clés d'identification est resté le même : il consiste à parcourir un graphe d'identification où chaque nœud est une question et où la réponse à la question ramène à un autre nœud et donc à une nouvelle question.
L'inconvénient de cette méthode est qu'on ne peut ni émettre d'incertitude ni laisser une question sans réponse. Cela pose donc un problème lorsque le caractère n'est pas observable.
L'apparition de méthodes d'identification assistée par ordinateur (IAO) avec le programme INTKEY [Dallwitz, 1974, site web de DELTA] a marqué un pas dans l'histoire de l'identification
Crédits
Le projet
Liens utiles
2
3
4
5

Fig. 2 – Pollen ID interface
(1) the menu, (2) Access to on-line identification services, (3) direct access to photos and videos, (4) details of the selected menu, (5) about Pollen ID.

The pollen ID project pays a special attention to beginners with the production of a rich information on palynology, from pollen extraction techniques to a glossary and interactive images and films for basic training.

3.1 Learning pollen morphology

In pollen ID, the user can explore easily all data: definitions, drawings, and images when necessary. The glossary has been inspired from [2]. In each page, hyperlinks are coloured and can redirect towards definitions and their associated drawings. Interactive drawings are managed by the rollover technique allowing the users to explore, to discover and to be familiarized with the terminology of pollen anatomy (Fig. 4). Then the beginner can learn the basic concepts of palynology.

The interface also combines real views (pollen and plant photos and pollen observation movies). The videos are constructed from a sequence of pictures (microscope X60), about 50 photos for each pollen, to have a good view of the total volume of pollen. The user can stop the movies as he wants, to compare with drawings describing anatomic structures. In the future, the ID project will intend to propose pollen photos with superimposed drawings in order to show the pollen characters directly on pollen views.

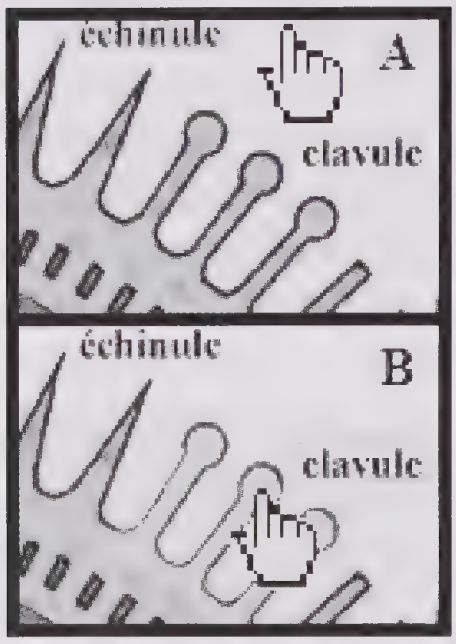


Fig. 3 – Example of an interactive drawing. The rollover technique highlights the morphological structure pointed by the user and displays the related definitions.

3.2 Pollen ID identification keys

Pollen ID includes at this step: two interactive free access keys, one dedicated to beginners and the other to advanced researchers. These two keys have been refined after [1]. They use the Xper^2 identification process and are a free access system available on-line. At the end of identification process, users can access the complete information on the taxon (Fig. 4).

3.3 Pollen ID Taxonomic forms

Pollen ID produces taxonomic forms that gather nomenclatural data, geographical information, textual pollen descriptions, pollen images and

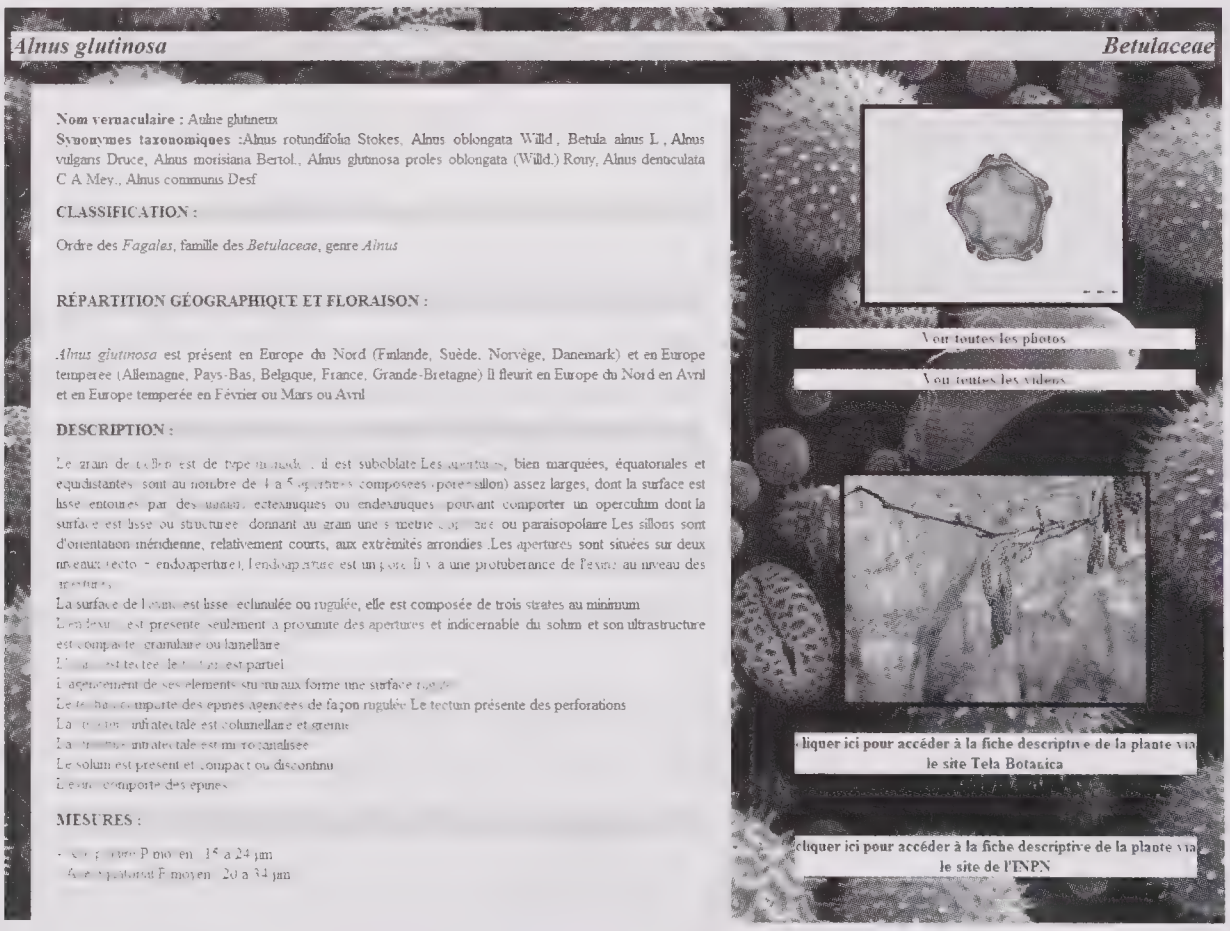

Alnus glutinosa
Betulaceae
Nom vernaculaire : Aulne glutineux
Synonymes taxonomiques :Alnus rotundifolia Stokes, Alnus oblongata Willd., Betula alnus L., Alnus vulgaris Druce, Alnus morisiana Bertol., Alnus glutinosa proles oblongata (Willd.) Rouy, Alnus denticulata C.A.Mey., Alnus communis Desf
CLASSIFICATION :
Ordre des Fagales, famille des Betulaceae, genre Alnus
RÉPARTITION GÉOGRAPHIQUE ET FLORAISON :
Alnus glutinosa est présent en Europe du Nord (Finlande, Suède, Norvège, Danemark) et en Europe tempérée (Allemagne, Pays-Bas, Belgique, France, Grande-Bretagne) Il fleurit en Europe du Nord en Avril et en Europe tempérée en Février ou Mars ou Avril
DESCRIPTION :
MESURES :
Voir toutes les photos
Voir toutes les vidéos
Cliquer ici pour accéder à la fiche descriptive de la plante via le site Tela Botanica
Cliquer ici pour accéder à la fiche descriptive de la plante via le site de l'INPN

Fig. 4 – Taxonomic form example. (1) textual description (the part "DESCRIPTION" is produced in natural language from the structured Xper^2 knowledge base), (2) pollen photos access, (3) movies access, (4) plant photos access, (5) external links. Some words are hyperlinked to definitions in the glossary.

movies, plants photos and links to external websites (Fig. 4). All textual pollen descriptions are automatically generated from the Xper^2 structured descriptive data with added hyperlinks to the pollen glossary. Thus, the user can go through the different items to have a look at all information, and all the parts of the website are consistent.

3.4 Taxonomic research

The user can also find directly a taxon by choosing it in the lists of forms, photos or movies. A classification is available with scientific and vernacular names. In the future the project will include an interface to construct diagrams from pollen inventories, and links with vegetation and climate data for environmental studies and archeo-paleo climatic reconstructions.

4 Conclusion

The Pollen ID project is presently restricted on the European and Mediterranean geographical area, but it will be enlarged to other regions as well. This project is still in progress; its content and user interface – presently in French - will be available soon in English. In its final shape, the Pollen ID project will include palynological applications such as pollen determination tests, several original pollen analysis exercises with representations in diagrams, and an easy vegetation and climate interpretation. Pollen ID is accessible at http://lis-upmc.snv.jussieu.fr/pollen/.

Acknowledgement

This work has been done with CNRS, MNHN and UPMC funds. We thank JP Cazet for pollen process, and pollen photos and movies production..

References

[1] J. Lebbe, S. Nilsson, J. Praglowski, R. Vignes and M. Hideux, “A microcomputer-aided method for identification of airborne pollen grains and spores”, *Grana*, vol. 26, pp. 223-229, 1987.

[2] W. Punt, P. P. Hoen, S. Blackmore, S. Nilsson and A. Le Thomas, “Glossary of pollen and spore terminology”, *Review of Palaeobotany and Palynology*, vol. 143 (1-2), pp. 1-8, 2007.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 389-393.
ISBN 978-88-8303-295-0. EUT, 2010.

The UK urban tree survey

Bob Press

Abstract — Public communication of science shares various educational aspects with more formal learning experiences but differences in both the aims and the target audiences make for subtly different approaches. The UK urban tree survey is an example of the requirements and possibilities of a public-orientated project relying on the use of identification tools.

Index Terms — trees, identification tools, survey, public communication of science.

1 Introduction

Depending on the target audience, projects with an educational aspect need to adopt different approaches to learning. Public communication of science differs in small but significant ways from projects exclusively aimed at formal learning in schools (Tab. 1).

Schools (formal learning)	Public (informal learning)
Not voluntary for participants	Entirely voluntary for participants
Linked to curriculum	Linked to personal interest
Narrow within-group ability range	Broad within-group ability range
Direct support available	Little/no direct support available
Participants use results	Participants do not use results

Tab. 1 – Considerations for formal versus informal learning.

As a project relying on the active participation of the general public, the UK tree survey (launched in July 2010) is an example of a wide-ranging educational activity but with a focus on public participation and demonstrates some of the ways in which these differences can be addressed.

2 Context

Why survey trees at all – especially in the UK which has one of the best known

Bob (J.R.) Press is with the Natural History Museum, Cromwell Road, London, SW7 5BD. E-mail: j.press@nhm.ac.uk.

and most recorded floras in the world? Trees are currently a focus of interest in the UK, following studies such as that on urban health and forestry [1], and investigations into the potential role of trees in areas such as carbon off-setting. The Department of the Environment, Fisheries and Agriculture (Defra) recently announced plans to plant 1 miilion trees across the UK.

In addition to the wild flora, there are copious (though patchy) data relating to trees managed by such organisations as the Forestry Commision and local authorities. Despite this, the urban forest remains relatively little known. In particular, data on trees in private gardens represent something of a black hole.

Against this background, the Natural History Museum in London decided to launch a web-based survey of urban trees. For a survey of this type public participation is vital since only members the public have access to many of the areas of interest – especially to private gardens. Since public participation in such surveys is entirely voluntary, it is intimately linked to personal interest. Previous experience (Bluebells survey, OPAL surveys) indicates a strong willingness by the public to take part in what they perceive as 'real science' as long as the scientific reasons behind the project are clear.

The scientific aims of the urban tree survey are:

1. Gaining a more precise understanding of the make-up of the urban forest e.g. the constituent species
2. Changes in tree demographics (and perhaps their causes)
3. Potential impact on (and changes in) other wildlife relying on trees
4. To gain phenological data and insights into the effects of climate change (changes in the seasons as indicated by flowering times, which species are now flourishing where in the UK etc).

Trees are highly visible within the urban environment and have well-known benefits for urban populations. The scientific aims of the project are easily seen as relevant to the public at large.

3 The urban tree survey

A major factor in designing the survey is that the public user has no on-the-spot expert to consult directly such as a botanist or teacher; essentially the user is working alone. For the identification and mapping tools in particular, clarity and ease of use are essential, with an obvious progression through the necessary steps. Clear, unequivocal instructions are needed at every major decision point lest the user lose interest and stop.

A pilot survey focusing exclusively on cherries was run from April to June 2010. This garnered extensive media attention at a national level and received huge support from the public. The experience was used to modify the main survey before it was launched in July 2010 (see 6. Lessons learned).

3.1 Species included in the survey

A total of 80 'trees' were included in the survey. Some of these are individual species, others groups of species and yet others are genera. This apparent complication is necessary for three reasons. 1) Given that we do not know

precisely which taxa occur in the areas to be surveyed, producing a definitive key to individual species is impossible. One of the reasons for the survey is to gain an initial idea of which taxa are present and then to refine the keys in the light of these data. 2) A key including every possibility would be too large and unwieldy. 3) There are some taxa e.g. *Sorbus* which are simply too difficult for non-experts. A reduction to groups enables users to cope with taxonomically difficult trees.

3.2 Identification tools

Prevailing wisdom is that keys for the public must be short, simple and entirely devoid of technical terminology. We know from *KeyToNature*'s work that this is not necessarily true for schools and it is no more true for the public. However, keys which are obscure or fussy are confusing and counter productive for non-experts, so presentation is a major consideration. The choices at each step need to be clear and it must be possible to retrace the steps to correct any wrong turning.

The key provided is interactive and uses simple illustrations, images and text. As each step is negotiated it recedes on the screen while the next step takes centre screen. This guides the user through the key while previous steps remain visible to aid backtracking. An interactive version for i-phones and a printable key are also available.

3.3 Factsheets

A key alone is insufficient for identification purposes when there is no expert present to confirm the result. Thus, each exit point from the key is linked to a fact sheet for that species, enabling the user to confirm their identification.

3.4 Mapping

Familiarity breeds confidence. We used Google Earth, a mapping tool already familiar to many people. Our system is a hybrid of aerial photographs and street maps. Offering easy access is important so locations can be identified by several methods: by post code; grid reference; latitude/ longitude; or by simply zooming in. The simple system encourages users to continue and helps to avoid errors when adding records.

3.5 Recording data

A tree is added to the map with a click of a mouse and the user is asked to record data relating to it. While sufficient data to make the survey worthwhile is required there are limitations on what the users can provide due to their lack of expertise. There may also be a lack of support or encouragement (school children have a teacher to prompt them to go one step further, the public have no such presence).

We restricted data to eight fields including date of the record, identification,

type of site (private garden, street, park etc) and size of tree. This is in line with the concept of keeping the effort required to achieve a result to a minimum. The fields use drop-downs to ensure consistency of data entry.

When dealing with the public, plant names probably cause more misunderstandings than any other factor. The public prefer to use vernacular names raising all manner of complications. A particular problem arose with cultivars, of which there are many hundreds. They are often sold under names such as *Prunus* 'Amanogawa' (= *Prunus serrulata*) or *Prunus* 'pandora' (= *P. yedoensis*). Some users were convinced these were species names and, instead of keying the tree out, simply believed their species had been omitted and did not enter a record.

We provided fields for both vernacular and scientific names, with an autofill function for the field not selected. Again, drop-down lists prevent 'false' names from being added.

There is an option to upload up to three images of the specimen which helps the survey team to monitor records. No free text was allowed in any field. This is a safeguard against individuals posting abuse on the web site.

4 Other support

Given that the target audience is working without direct support, it is important to provide as much additional information as possible. The web site offers tips on tree identification, an on-line identification forum, glossary and references to other identification tools. There is an entire section devoted to learning support for schools wishing to use the survey.

5 Results

To date (July 2010), over 5000 trees have been recorded and mapped [2]. There have been more than 100,000 web site visits, with a bounce rate of 25%.

This is a baseline survey and one in its early stages. Statistical validity considerations aside, we are only able to make a generalised analysis of results for now. However, from the pilot survey we already have a snap-shot of *Prunus* species in the UK urban forest, with some ideas of which species are most frequent in different site types, which are most widespread and so on.

Providing feedback is vital to maintain the connection with the audience. The public do not themselves use the results of surveys such as these but they do want to see what difference their efforts have made. The initial results of the cherry tree survey have already been posted on the web site.

6 Lessons learned

The pilot project was informative both in terms of the data gathered and in the challenges of conducting a survey of this kind. The positive lessons learned are:

1. The public are keen to participate – as long as they can be attracted by the project (it is voluntary)
2. The public can cope with quite sophisticated methods

3. The project must be linked to a real scientific question/investigation and tie in with personal interest rather than a curriculum
4. Sufficient data can be collected
5. Unexpected data appeared e.g. on harvesting and cultural practices associated with cherries
6. Contacts with other groups were made (e.g. The Orchards Initiative, local authorities, tree warden groups)

There are also negatives:

1. Data accuracy and data usability - since the public do not actualy use the data, they may be less interested in its accuracy
2. Verification of data – a major factor but one we must live with. Only the recorders have access to the sites although posting images is of considerable help in weeding out wrong or false records
3. Technical problems e.g. Explorer initially failed to cope with the number of records on the map. This and the point above relate to lack of support or how we provide support
4. Such projects risk continually re-inventing the wheel
5. There is no clear system for migrating data upwards/outwards to other organisations and potential users e.g. GBIF

7 Next steps

The cherry tree survey has now been subsumed within the similar but larger project covering all urban trees in the UK [3]. Both surveys will be refined in light of the results received by the end of 2010 and repeated for two more years. All data will be made available to other users.

8 Conclusion

Surveys based on identification of organisms can be remarkably successful on two fronts: in gathering scientific data and in communicating science to the public. As with any such project, careful consideration of the audience's needs are paramount. Clear, consistent presentation of information (including the reasons behind the tasks) and using familiar technology are also requirements.

Acknowledgement

The author wishes to thank Kate Evans, Claire Gilby, Sam Rae, Sheila Sang, Mike Sadka, and Philippa Watson for contributions to preparing and launching the web site. This work was supported in part by a grant from the Gulbenkian Foundation.

References

[1] Liz O'Brien, Kathryn Williams and Amy Stewart, "Urban health and health inequalities and the role of urban forestry in Britain: A review". *Forest Research*, 2010.

[2] http://www.nhm.ac.uk/nature-online/british-natural-history/urban-tree-survey/results-map/index.php, 2010.

[3] http://www.nhm.ac.uk/nature-online/british-natural-history/urban-tree-survey/index.html, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 395-400.
ISBN 978-88-8303-295-0. EUT, 2010.

Tree School – A new Innovation for Science and Education

Della Hopkins, Karen James

Abstract — The current decline in the popularity of science at school has resulted in many current activities to encourage and enthuse children in these subjects. One such initiative has been born from a partnership between the Natural History Museum, London, and the Cothill Educational Trust, which aims to engage children in current biodiversity science. This project, named Tree School, combines traditional methods used in botanical taxonomy with new research activities using DNA barcoding to investigate local biodiversity, and the applications of these techniques on a larger scale. The challenges faced and potential opportunities realised have been investigated with a series of pilot workshops, in preparation for the two-year operational phase in 2011-12.

Index Terms — DNA barcoding, identification, trees.

1 Introduction

The current decline in popularity of the science subjects for school age children has been recognised both in the education sector and also by scientists. With fewer children choosing to continue to study science, there may be a shortage of scientifically literate and trained professionals in the future. A recent Research Councils report highlights this predicament, encouraging the development of innovative initiatives between scientific researchers and schools [1].

Scientists at The Natural History Museum (NHM) in London were keen to address this problem, and began to develop ideas into how modern scientific techniques could be effectively communicated to school children. Public engagement and learning already play a huge role within the Museum; extending these principals through the scientific researchers was an obvious development.

Simultaneously, the Cothill Education Trust was considering ways to excite and encourage the children at their schools in science. A partnership was formed between the NHM and the Cothill Trust, and a new initiative developed with the aim of enabling current scientific practices to be understood and undertaken by

D. Hopkins is with the Natural History Museum, London, SW7 5BD. E-mail: d.hopkins@nhm.ac.uk.

children. The Tree School project forms a continuing collaboration between the two parties, to encourage a greater interest in science in children.

2 The Tree School

The scientific aim of Tree School was to design and pilot ways of involving children and other non-experts in biodiversity science, specifically international DNA barcoding campaigns. Pupils from the Cothill Educational Trust schools, as well as others from schools across the country, collaborate with NHM researchers to collect scientifically relevant data in this emerging scientific field. This is achieved by attending a five-day workshop at the Old Malthouse School, a fully-equipped educational centre in Dorset, which combines traditional classroom, practical and outdoor teaching techniques. The Old Malthouse accommodates the pupils, teachers and scientists comfortably, allowing a relaxed atmosphere for both study and leisure time.

The workshop incorporates the more traditional methods used in taxonomy and identification with modern laboratory-based work. This enables a deeper understanding of the subject, and also demonstrates how different research overlaps and compliments the other. The trees situated on the Old Malthouse grounds are being used, with the children investigating different techniques to identify the trees to species level.

2.1 Taxonomy

Field work is carried out at the beginning of the week, where the tried and tested methods still used by botanists in the field today are demonstrated. Each pupil chooses a tree, and captures the relevant data, such as a GPS point, relevant observations and the correct species name. Identifications are made using a variety of binomial keys and pictures, namely:

KeyToNature: Key to trees at the Old Malthouse [2],
The Field Studies Council: 'Tree Name Trail' key to common trees [3],
The Woodland Trust: Leaf identification swatch book [4],
The Collins Tree Guide [5].

A specimen from the tree is also taken and pressed, which is later made into a herbarium sheet incorporating all of the gathered information.

2.2 DNA Barcoding

Leaf samples are also taken from the trees for use in DNA barcoding. This is a compelling new tool, promising dramatic improvements in the rate and accuracy of biodiversity inventories, and the potential to identify samples to a species level. Within the fully-equipped laboratory, the children carry out DNA extraction and PCR amplification and evaluation for their own specimens. These extracts are then taken to the NHM sequencing laboratory, and the DNA sequences produced. The results are added to TreeBOLD [6], the international barcode of life data system, where they will be available to the scientific community.

3 Pilot Workshops

In 2011, a series of pilot workshops were undertaken to develop these ideas, concentrating on the suitability of the techniques and the resources used. Opinion has been sought from the teachers, children and scientific researchers attending these workshops, with the findings and lessons learnt used to finalise preparations of a series of workshops to take place during 2011-2012. It is anticipated that pupils from both private and state schools will attend these workshops.

A series of questionnaires were carried out both before and after attending Tree School for the children and the teachers. Thirty five children aged between eleven and twelve were questioned, from two different schools and currently in year 7 of the UK system. A series of background questions were asked upon their arrival, including their favourite subject, whether they enjoyed science at school, whether they would like to work in science when they are older, and their perception of a scientist. Three teachers present for the entirety of the course participated in the survey, comparing their perceptions of Tree School with their findings having attended.

3.1 Teachers

The common concern prior to Tree School was that the work would be too advanced for the age group attending. Conversely, each participant stated that any opportunity for students to broaden their horizons can only be a positive experience. The teachers also expressed a frustration that there is little opportunity for them to explore new aspects of science, or develop children's areas of interest, due to syllabus constraints. Most work taught during term-time is exam-driven, which of course is a necessity.

At the end of the week, the teachers were pleasantly surprised at the amount the children had understood, and their level of engagement in both classroom lessons and practical sessions.

3.2 Pupils

Upon arrival at Tree School, the most popular subject proved to be French, with around one third of the children naming it as their favourite subject. The next most popular subjects were History, Geography and Art, followed by Science with four votes. English, Maths and Latin were also mentioned. The majority of the children stated that they enjoy science at school, although very few have considered continuing with science in the future. Most of the group were unsure about what they wanted to do as a career.

An interesting observation is the high number of children who chose French as their favourite subject. The Cothill Trust has one school in France, the Château de Sauveterre, where the children spend one term learning French intensively to a much greater extent than can be covered in normal lesson time. This has obviously had a positive effect on their enjoyment of the subject. It can therefore be assumed that with extra insight and learning into a subject, children

can easily be enthused and encouraged to pursue a subject in the future. With 72% of the children already enjoying science at school, but the same number unsure as to what they would like to do as a career, it is possible that more children could choose to study science at a higher level if given the opportunity to engage in real science at a younger age.

3.3 Scientists

The workshops also proved an enlightening experience for the research scientists from the NHM. It was felt that the information was communicated effectively, supported by informal reports back from the teachers, students and even the students' parents. The level of engagement with the children grew over the week, leading to some thought provoking discussions. It provided a new perspective on the subject, as children ask very different questions to those usually posed by peers and other academics, and resulted in a two-way enthusiasm between the scientists and the students.

It was also interesting to see how the children reacted to real-life experimental problems faced by scientists. Schools usually teach canned experiments with known outcomes, so pupils are usually able to achieve the correct result, or know where they may have made a mistake. Often, a species may not appear in a guide book, which encourages the child to think about how they may be able to identify their tree. Also, the processes used to extract and amplify the DNA can often fail, for no obvious reason, which can often happen.

4 The Workload

It is obviously imperative to ensure that the aims of the Tree School are fully met for each group of children attending the workshops. It is important that they end the week with a proven increased knowledge of identification and scientific methodologies, and also hopefully an increased enthusiasm for botanical science.

Much of this can be gauged from the question and answer sessions during classroom and outdoor activities. In addition to queries for clarification, inquisitive questioning increased and many sensible further and leading questions were asked as the week progressed. This was particularly exciting to see, as many of the children became engaged in the project and were inspired to find out additional information.

Feedback from the children was important in ascertaining the work load, and level of understanding of the classroom aspects. All attending pupils were questioned on the level of the work undertaken – was it too easy, about right, or too difficult. Over 90% of the children claimed the work was about right, stating that although much of it sounded complicated, the methods were easy to carry out with well-explained instructions and help. Three participants thought the work was too easy, and none that it was too hard. Whilst this is a positive result, it is important to consider preparing some additional information for those pupils wishing to investigate the subject further.

Direct testing was also carried out at the end of the week in the form of a

tree identification quiz. A selection of trees was labelled, and the children were given the task of identifying them using the resources and techniques given to them previously. The results were encouraging, with many achieving full marks. Again, the children were questioned about this task, and the different keys available to them. The *KeyToNature* key, developed especially for the Tree School project, was a clear favourite with 23 of the 35 children choosing this as the most user-friendly resource.

5 Opportunities and Challenges

There is a huge opportunity to develop further projects using the same template at the Old Malthouse. The school is a short distance from the Jurassic coast, affording opportunities to study geological formations and other geographic principals. Additional biological courses could also be offered, for example studies into the invertebrate communities of the site, associated with different habitats or plant species. Many of these ideas could be developed in conjunction with other departments within the NHM or could be provided by expertise available directly to the Cothill Trust.

Whilst the opportunity to develop the project at the Old Malthouse is an exciting and achievable prospect, the possibilities of extending the format on a wider scale, even internationally, is not straightforward. Whilst the concept is easily expanded, the varying requirements of different educational structures must be established fully before undertaking the work. This will involve working closely with the key educators to determine prior knowledge, as curriculum may vary greatly in different countries.

6 The Bigger Scientific Picture

Whilst the Tree School project aims to provide scientific skills to children, and engage them in cutting-edge science, the project also is invaluable to current scientific research. Currently, an international consortium of scientists are working to build libraries of short, specimen-linked DNA sequences against which unknown specimens might be identified to the species level. This will only succeed as an identification tool for unknown specimens if a comprehensive library of DNA barcodes is first developed. All DNA sequences generated by the children will be entered into this database.

It is also thought that heavy sampling can improve the overall efficacy of a DNA barcode database, and will help to test the expectation of low intraspecific variation as compared to interspecific variation. If sampled intensively as part of this project, the British tree flora could serve as a test-bed for the effects of intensive specimen sampling on DNA barcode performance.

Acknowledgement

The authors wish to thank the Cothill Educational Trust for their support and collaboration with this project.

References

[1] Anon. Research Councils UK: Engaging Young People with Cutting Edge Research: a guide for researchers and teachers. www.rcuk.ac.uk/per, 2010.

[2] P L. Nimis, B. Press and S. Martellos, *KeyToNature*: Key to trees at the Old Malthouse. http://dbiodbs.units.it/carso/chiavi_pub21?sc=446, 2010.

[3] J. Oldham and C. Roberts, *The tree name trail: A key to common trees* (2nd edition). Field Studies Council / Forestry Commission. FSC Publications, Shrewsbury, Shropshire, 2003.

[4] The Woodland Trust: Leaf identification swatch book.

[5] O. Johnson and D. More, *Collins Tree Guide*. Collins, UK, 2006.

[6] TreeBOLD: http://www.boldsystems.org/views/login.php, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 401-404.
ISBN 978-88-8303-295-0. EUT, 2010.

Engaging Schools in Cutting Edge Science: From the Educator's Perspective

Adrian Richardson, Della Hopkins

Abstract — The field of scientific research, by definition, is constantly developing new techniques and adapting current thinking in order to address pertinent issues. With curriculum constraints and exam-based teaching, it is becoming increasingly challenging to engage young people in new ideas and methods, and thus facilitate them in becoming the scientists of the future. A new project developed though collaboration between the Cothill Educational Trust and The Natural History Museum aims to develop a deeper understanding of biodiversity science in pre-GCSE aged children, kindling a real excitement for the science subjects at school.

Index Terms — education, engaging and enthusing, project development, scientific research.

1 Introduction

A recent government report has highlighted the current predicament faced by science and education. Whilst it is accepted that science and scientists are crucial globally at economic, environmental and social levels, less children are studying science subjects at school. This is especially true at GCSE and higher levels, where children choose their subjects. The danger is that if this trend continues, there will be a shortage of science professionals in the future [1]. It is therefore imperative that children have the opportunity to engage in science at an early age; to enthuse and inspire them to choose to continue to study science, and to realise the career opportunities available to them.

This view is shared within the scientific community, with an awareness that students are often unaware of current methods routinely used by scientists. "Exciting new areas of science typically do not appear in science classrooms and textbooks until many years after their inception. The result is that too many students are never afforded opportunities to learn about the cutting-edge discoveries that make biology so exciting to professional scientists [2].

A. Richardson is Principal of the Cothill Educational Trust, Abingdon, Oxfordshire, OX13 6JL. E-mail: adr@cothill.net.
D. Hopkins is the project manager of Tree School on behalf of the Natural History Museum, London, SW7 5BD. E-mail: d.hopkins@nhm.ac.uk.

2 The Partnership

In order to address these problems, a partnership was formed between two like-minded parties: the Cothill Educational Trust [3] and The Natural History Museum of London (NHM). Through a shared interest in combining research, learning and public engagement, a project began to develop to bring together educators and scientists.

2.1 The Educators

The Cothill Educational Trust was established in 1967 by the Trustees of Cothill House, a boys' boarding school, with the aim of providing first class education to children up to the age of 14. The Trust now incorporates seven schools, including the Château de Sauveterre in France, and the Old Malthouse, a science education centre used for the new collaboration between the Cothill Trust and the NHM.

2.2 The Scientists

The Natural History Museum is a well-known and respected organisation. Not only does it house internationally-respected scientists, but the association provides the project with an identifiable name within the public realm. A collection of scientists from the NHM are collaborating on the project, with a range of specialist expertise, including molecular genetics, botanical taxonomy and biodiversity conservation.

3 Needs Identified and Aims Agreed

In early 2009 a series of meetings were held to explore the potential of the project and agree the aims between the two partners. These meetings not only included representatives from the Cothill Trust and NHM scientists, but also involved staff from other NHM departments including Learning and Interactive media.

The needs and requirements of both parties were identified. Whilst the Cothill Trust required the support of the museum, both in terms of scientific expertise and project management, the NHM needed the experience of an educational partner and access to a suitable place in which to carry out the teaching. It was also imperative for the museum to involve suitable staff for the teaching aspect of the project. It required enthusiastic and personable presenters who can interact with children, but who are deemed specialists in order to provide "cutting edge" science and thus a premium "product".

Once the needs and provisional capacity of each partner had been determined, the combined project aims and deliverables could be decided. It was agreed that the NHM - Cothill Educational Trust Project 'Tree School' would establish proof-of-concept in joining scientific research objectives together with science education imperatives through botany and DNA barcoding. Specifically:

To design, pilot, optimise and communicate methods for involving schoolchildren

and other non-experts in international DNA barcoding campaigns.

To promote the development of a scientifically and environmentally literate citizenry.

To increase the scale on which biodiversity science can be undertaken.

A start-up phase was established in order to develop the relationship between Cothill and NHM, with the Cothill Educational Trust providing infrastructure, equipment, logistics, teachers, pupils and the Natural History Museum contributing the science and learning.

4 The Role of the Cothill Educational Trust

4.1 Location

The Old Malthouse, situated on the Isle of Purbeck, Dorset, has been developed into a field centre fully equipped to provide five-day residential courses for up to 32 children at any one time. The Old Malthouse was a boarding preparatory school, but closed in 2007. It has been completely refurbished, with dormitories for the children and individual rooms for the teachers and visiting scientists. The concept when redesigning the interior was to provide a safe and comfortable environment for all participants, to create a relaxed atmosphere for enhanced learning and enjoyment. The classroom blocks were also updated to create a laboratory for the DNA barcoding, and a herbarium area for the storage of specimens.

4.2 Schools

The boarding and day schools managed by the Cothill Trust will be amongst the first schools participating in Tree School. Attendance will also be extended to state schools with links to Cothill, funded by charitable donations to the Trust. Following planned publicity, it is expected that schools will pay to attend Tree School, although state schools will continue to be subsidised by the trust.

The importance of the accompanying teaching staff can not be underestimated. They will maintain overall control of the classroom, in terms of discipline, but can also act as an intermediary between the children and the scientists. They can refer to recent lessons, to demonstrate how topics interact and overlap, and can also ask leading questions of their own to engage with the scientists.

5 Challenges and Benefits

The initial phase of this project has seen a series of three trial workshops through the summer of 2010. The scientists have worked in collaboration with three of the Cothill Trust schools in order to ensure the smooth running of the project during 2011-2012.

5.1 The Challenges

The principal challenge faced at the onset of the project was to find a suitable location, which was able to provide accommodation, laboratory space and enough outdoor space and suitable 'field' locations. This need also brings with it a huge financial requirement, not only for the initial set up, but also for the continuing maintenance and upkeep of the site. The recruitment of enthusiastic scientists from the NHM proved to be straightforward, and a high demand from the teachers and pupils was established.

Modifications have been made between workshops, in order to pitch the science at the right level for the group attending. Timetable alterations have been needed, due to varying group sizes and unpredictable weather conditions.

5.2 Potential Benefits

The immediate benefits of Tree School are apparent, with the provision of learning for an increasing number of children.

The concept also has the potential for development, both for children and adults. A workshop specifically for teachers could provide ideas and training, and incorporate Continuing Professional Development opportunities. Interest has also been expressed by other non-scientific adults, which could perhaps be developed as a summer school event. The project design can also be applied to further scientific projects, and other academic subjects, all with the principal aim of engaging young people in the excitement of cutting edge research.

6 Conclusion

This endeavour is a marked change from the 'canned' experiments schools are required to provide during scientific learning. Whilst these lessons are an important aspect of understanding science, the inclusion of current, innovative scientific principals and experiments allow an insight into science and research, as well as an opportunity for the children to challenge their questioning skills and feed their desire for learning. It is important to fire an interest in science at a young age in order to secure the scientists of the future.

Following the pilot workshops, a continued enthusiasm for the subject has been reported once back at school. Many of the methods used at Tree School can be replicated away from the field centre, such as trying to identify and map the trees on the school premises. It is planned that once schools become engaged with this project, and have a long-term relationship with the NHM, ideas will be developed and lead to further collaborative projects.

References

[1] Anon. *Nurturing tomorrow's scientists*, Department for children, schools and families. DCSF Publications, Nottingham, 2007.

[2] A. Jurkowski, A. H. Reid and J. B. Labov, From the National Academies: Metagenomics: A Call for Bringing a New Science into the Classroom (While It's Still New). Center for Education, National Research Council, Washington, DC. *CBE—Life Sciences Education*. vol. 6, 260 -265, Winter, 2007.

[3] The Cothill Educational Trust. http://www.cothill-trust.demon.co.uk/index.html, 2010.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 405-409.
ISBN 978-88-8303-295-0. EUT, 2010.

Educational or emotional languages? An interactive experiment with the Lucanian flora (S-Italy)

Riccardo Guarino, Patrizia Menegoni, Sandro Pignatti

Abstract — In the frame of dissemination activities for a still-in-progress work on the Sites of Community Importance (see EU Directive 92/43) of Basilicata (a region of Southern Italy), an interactive tool (IIT) for the identification of vascular plants growing there has been illustrated to two groups of people, following two different approaches: one focused on textual parts and on scientific accuracy, the other on images and on the visual comparison of different objects. The reactions were measured in terms of number of accesses to the IIT, elapsed time from the demonstration to the first individual access, and number of queries in the first week after the IIT was distributed. The most clicked options were recorded as well. People who followed the emotional/visual approach proved to be significantly more interested in the IIT than those who followed the descriptive/scientific approach. It seems that to raise the interest of non-experts to the identification of plant species and, more in general, to the study of biodiversity, words should be kept at minimum, while the quality of the images and their "appeal" are essential.

Index Terms — Basilicata, flora, interactive identification tools, people's reaction.

1 Introduction

There is general agreement that for scientists it is important to foster public knowledge on biodiversity and ecosystem's functioning. However, all too often educators think about this focus in a fragmented manner, either as an important end in itself, or as a contribution for enhancing people's awareness on their responsibility towards nature and on the effects of human impact. In the first case, a classical academic approach is followed and often

R. Guarino is with the Dept. of Botanical Sciences, University of Palermo, I-90123. E-mail: riccardo.guarino@unipa.it.
P. Menegoni is with E.N.E.A. - C.R. Casaccia, Via Anguillarese 301, S. Maria di Galeria (Rome), I-00123. E-mail: patrizia.menegoni@casaccia.enea.it.
S. Pignatti is with the Dept. of Plant Biology, University of Rome "La Sapienza", I-00165. E-mail: sandro.pignatti@gmail.com.

the efforts towards popularization are limited to the simplification of concepts and to a drastic reduction of the provided information. Nature and biodiversity tend to be depicted as a special selection of vertebrates and big, colourful invertebrates that sometimes interact with the most attractive plants growing in a given place, neglecting a myriad of other living organisms. In the second case, a paternalistic approach is followed: the few who know provide strong evidence that the survival of a relevant percentage of living organisms is at risk, planning informative campaigns on most striking examples (polar bears, coral reefs, tropical rainforests...), following the theory that what does not raise people's interest has no value. The most typical, although not very logical, conclusion of these campaigns is that humans should respect any form of life not only for ethical reasons, but also because preserving the integrity of natural ecosystems is an essential need for the survival of ourselves.

We think that to foster people's knowledge on biodiversity and ecosystem's functioning is important *per se*, but that the success of these non-academic outcomes can be better achieved through the development of new social and emotional learning techniques, which do not necessarily imply excessive simplification and reduction of concepts and information.

In order to test what languages and type of information best stimulate people's response and intellectual behaviour, a simple experiment has been carried out by means of an interactive identification tool on a regional flora. The results are presented here.

2 Material and methods

An interactive identification tool (IIT) on the vascular flora of Basilicata (a region of Southern Italy) has been presented and distributed to 54 people attending a seminar on the Sites of Community Importance (see EU Directive 92/43) of the same region.

The IIT was an excerpt of the digital utilities developed for the second edition of Pignatti's *Flora d'Italia* [1]. Core of the IIT is a multi-entry key where users can filter the species by making their own choices in a set of non-hierarchized fields and options (see [1], Fig. 2). The template of the multi-entry key consists of four main components: a number of fields, a list of options for each field, a scroll-down table describing and illustrating every single option, a directory command leading to the filtered species.

For the presentation of the IIT, attendants were divided into two groups of 27 persons each. They were working in a computer lab of the University of Palermo. Both groups consisted of post-graduates with similar age and instruction.

Two slightly different versions of the IIT were presented: in the first one, the scroll-down table of the template of the multi-entry key was describing *but not* illustrating the single options, and the directory command was leading to a classical dichotomous key; in the second one, the scroll-down table was describing *and* illustrating the single options, and the directory command was leading to a panel where the identification of a species took place through the visual comparison of different images and, *after that*, through the (optional) reading of the diagnostic characters.

The "style" adopted during the presentation of the IIT was also different: in the first case, more emphasis was given to the scientific accuracy of the information provided, and the identification of the specimens selected for the experiment was carried out as an individual activity (each participant had his own computer and specimen); in the second case, more emphasis was given to the images, and the identification was carried out as a group activity, with discussions on the available options and plenty of jokes on the visual skills of the people involved.

People's reactions were measured in terms of number of accesses to the IIT, time elapsed from the demonstration to the first individual access, number of queries in the first week after the IIT was distributed. The most clicked options were recorded, as well. The significance of the differences observed in the two groups were checked with Student's t test.

3 Results

People who followed the visual/participatory approach (group 2) seemed to be significantly more interested to the IIT than those who followed the descriptive/ individual approach (group 1), at least concerning the number of accesses to the IIT in the first week and the time elapsed from the demonstration to the first individual access. The number of queries, i.e. the number of options (out of 1496 possible ones) experimented by each user was not significantly different (see Tab. 1 for details) within the two groups, nor were the most clicked options.

The most clicked options pertained to the following fields: "regional distribution", "life form", "group" (a field including 12 options based on simple floral characters, like symmetry and number of floral parts), "colour of the flower", "veining of the leaves".

	Mean	*St. Dev.*	*0,05 C.I.*
Nr. of accesses - group 1	9.04	5.54	2.09
Nr. of accesses - group 2	13.87	5.22	1.97
Elapsed time (hrs) - gr. 1	87.70	38.99	14.71
Elapsed time (hrs) - gr. 2	62.44	42.88	16.17
Nr. of queries - group 1	212.07	127.56	48.12
Nr. of queries - group 2	245.90	115.82	43.69
	Nr. acc.	*El. time*	*Queries*
Pearson's correlation	-0.0501	0.0675	-0.1298
Stat t	3.1791	-2.5447	0.9597
P(T≤t) - one tail	**0.0020**	**0.0087**	0.1732
P(T≤t) - two tails	**0.0039**	**0.0175**	0.3464

Tab. 1 – Responses of the two groups to the considered parameters and their significance.

4 Discussion

Two relevant facts influenced users' behaviour in our experiment: the use of images and the presentation of the IIT as a kind of "social game".

In the academic communication and in the related formative activities, descriptive models are still largely based on textual forms and the learning process is all too often seen as the result of individual efforts.

Attractive colours and images, integrated with the innovative tools made available by information technology, can create a supportive environment where experiential activities can be carried out with a social and emotional involvement. This will more easily convey a durable acquisition of knowledge [2].

A community approach is vital in a learning process. Many recreational groups with online forums are already fostering the botanical culture of people. Some meaningful examples are, for Italy country: GIROS (www.giros.it), Acta Plantarum (www.actaplantarum.org), Flora delle Alpi Marittime (www.floramarittime.it), F.A.B. (www.floralpinabergamasca.net), G.M.Lu (gmlu.wordpress.com), Natura Mediterranea (www.naturamediterraneo.com), Botanica Italiana (www.botanicaitaliana.it). Each of these websites counts thousands of visitors and relies on a permanent virtual community with hundreds of supporters.

The sharing of images and experience enhances the individual learning attitude. The identification and characterization of species becomes a participatory process to which everyone can contribute with images, observations, new findings and, finally, with the correct identification of the diagnostic traits of a given species.

This process is complex and operates at multiple levels. A good IIT should be well-calibrated on different level of fruition: the availability of information must be easy, and contents have to be interesting for the whole community, from beginners to experienced scientists. For these reasons, the starting questions when implementing an IIT should be: "Which kind of user do we address? What information users are looking for? Where/how do they expect to find it?". The answers to these questions can help in designing a gradual availability of the contents, able to raise the interest of multi-level users, with no need to sacrifice the completeness of the information for ensuring better usability [3].

Italo Calvino said that, in order to be effective, information must be: *light*, *short*, *exact*, *visible*, *coherent* [4]. If applied to an IIT, Calvino's sentence means that the functions of multimedia objects must be perceived as simple and immediate by the user (*light*); assets and files should occupy a few bytes, in order to be loaded and recalled very quickly (*short*); textual parts should be essential, precise and organized in small blocks, with keywords and concepts well highlighted and illustrated (*exact*); the hierarchy of the fields and options to filter the species, as well as the whole structure of functions, commands and graphic templates should be evident (*visible*). Finally, the sense of innovation stays in the ability of creating harmonic and complete communication paths through the integration of heterogeneous components, keeping at the same time the unitary consistence of a project (*coherent*).

5 Conclusion

The commonest way to cultivate a hobby, or a scientific interest, is to share it with other people. Universities, schools, associations are social places where learning is, intrinsically, a social process. Members of a community do not learn alone, but rather in collaboration with their teachers, in the company of their peers, and with the support of their friends.

Emotions can facilitate or hamper the learning process and the ultimate success in the amount of knowledge acquired. Because social and emotional factors play such an important role, interactive tools aimed at popularizing scientific knowledge will be most successful when they integrate efforts to promote people's academic, social, and emotional learning.

Our experiment suggests that, in order to raise the interest of non-experts in the identification of plant species, the learning object must be visually attractive and the learning process must be "blended": words should be well calibrated with illustrations, concepts must be clear and essential. The quality of the images and their "appeal" are essential, as well as the possibility for users to interact online, to share information and contents with other users and with scientists, to become members of a botanical virtual forum that keeps active thanks to the inputs of a large number of people.

References

[1] R. Guarino, S. Addamiano, M. La Rosa and S. Pignatti, *"Flora Italiana Digitale": an interactive identification tool for the Flora of Italy.* In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 157-162, 2010.

[2] N. M. Haynes, M. Ben-Avie and J. Ensign (eds.), *How social and emotional development add up: Getting results in math and science education*. Teachers College Press, New York, 2003.

[3] R. E. Mayer, *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, 2005.

[4] I. Calvino, *Lezioni americane. Sei proposte per il prossimo millennio.* Oscar Mondadori, Milano, 1993.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 411-416.
ISBN 978-88-8303-295-0. EUT, 2010.

Online sharing educational content on biodiversity topics: a case study from organic agriculture and agroecology

Nikos Palavitsinis, Nikos Manouselis, Kostas Kastrantas, John Stoitsis, Xenofon Tsilibaris

Abstract — E-Learning Technologies and Standards are emerging as the dominant way to make educational content widely available. Approaches to these technologies should be domain-independent and easily adaptable to different contexts. Organic.Edunet aims at making content on Organic Agriculture and Agroecology widely available through a single point of reference. To achieve this, the project has adopted and adapted Open Software solutions and has built upon them to offer the Organic.Edunet Web Federation Portal and the Repository Suite of Tools. This paper presents the tools that were developed in the frame of Organic.Edunet project, serving as a guide for all individuals that aim at establishing similar tools in a field such as the biodiversity.

Index Terms — repositories, education, open access, metadata.

1 Introduction

The Organic.Edunet project aims to facilitate access, usage and exploitation of digital educational content related to Organic Agriculture (OA) and Agroecology (AE). From the technical viewpoint of the project's objectives, Organic.Edunet aims to support stakeholders producing content about OA & AE in order to publish it in an online federation of learning repositories and describe it according to multilingual, standard-complying metadata. This objective is accomplished through the deployment of the Repository Suite of Tools.

Also on the same basis, the project has deployed a multilingual online environment (the Organic.Edunet Web portal) that facilitates end-users' search, retrieval, access and use of the content in the learning repositories. Both tools deployed (the Repository Tool and the Web Portal) are already running smoothly on the web while

All authors are with the Greek Research & Technology Network, Athens, Greece. E-mail: palavitsinis@grnet.gr, nikosm@grnet.gr, kkastrad@grnet.gr, stoitsis@ieee.org, xtsili@admin.grnet.gr.

small changes are being made to ensure a smooth operation. Having completed the biggest part of the work involved, the Organic.Edunet partners have a clear view of all the complexity, problems, issues and challenges that were faced during the deployment of the tools.

This paper aims to briefly describe the tools produced in terms of their main characteristics and to present a part of the work that was carried out for launching them successfully. More specifically, the metadata application profile that was used to deploy the Repository Tools will be described, and the main parts of the Organic. Edunet Web Portal will also be analyzed. Overall, this proposal aims to demonstrate a complete process of making educational content available online through the use of e-learning technologies and standards. This paper also aims to serve as a reference point for ongoing or future projects that will deal with the issue of making educational content available, regardless the application domain. The proposed methodology and tools can be deployed in order to support education on biodiversity.

2 Tools and project results

2.1 Application profile

The IEEE LOM standard has been chosen as the basis for the metadata application profile to be used in Organic.Edunet. The schema is therefore termed as Organic.Edunet Application Profile (AP). It adopts many of the elements of LOM, specializing several of them in order to better describe learning resources on organic agriculture and agroecology. In each one of the nine (9) categories of LOM elements, a number of elements have been refined, in order to be used in Organic.Edunet [1].

2.2 Repository tool

The Organic.Edunet AP is implemented in a repository management tool that is being used for the organisation and management of learning objects. The repository tool can be used by content providers to create learning resources, collections, and complex learning objects such as online exhibits and educational paths. Moreover, the tool allows content providers to design and integrate educational templates that implement specific educational paths and digital exhibits. Using the repository tool, content providers can generate complex learning objects by reusing existing learning objects in their repository.

The repository tool is used to describe resources with appropriate metadata, and to publish resources in their own learning repository. When the resources are published in the individual repositories, they can be harvested from the Organic.Edunet Portal. When a resource is harvested, it is published on the portal where users can access it.

The repository tool allows content providers to connect their repository to the Organic.Edunet federation of repositories. In addition, the tool supports the connection of the repository with other federations of learning repositories. This is achieved through the adoption of open standards and specifications for the

exchange of search queries and the harvesting of metadata. These standards and specifications include the Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH, http://www.openarchives.org) and the Simple Query Interface (SQI) [2].

2.3 The Portal

The Organic.Edunet Web portal (http://portal.organic-edunet.eu) aims to facilitate access, usage and exploitation of digital educational content related to Organic Agriculture and Agroecology. It is a multilingual online federation of learning repositories, populated with quality content from various content producers. Its main purpose is to facilitate end-users' search, retrieval, access and use of the content in the learning repositories.

The development of the Organic.Edunet Web portal addresses one of the main objectives of the project. That is to integrate and specialize state-of-art technologies of the World Wide Web, in order to provide end-users with a single European reference point that will offer advanced services such as ontology-based searching and social recommendation and facilitate search, retrieval and use of the collected content. The global architecture of Organic.Edunet is thus composed of 2 major subsystems (shown in Fig. 1):

- **Organic.Edunet Web portal (top part):** It provides services to the portal users (such as learners, teachers, professionals, etc.), ranging from searching through the content in various ways (different search functionalities), to tagging of resources and bookmarking of resources in a personal online space. The users are able to see the resources and their metadata because this information is fed into the portal from the federation of repositories.
- **Federation of Repositories (lower part):** This subsystem includes the repositories containing either the learning resources uploaded by the content providers of the project, or their metadata or both.

The potential users of the Organic.Edunet Web Portal include: teachers, students, pupils, researchers, OA&AE professionals, general public, etc. The Web Portal is internally made of four components: the semantic repository Ont-Space, the semantic navigation module, the portal infrastructure and the social navigation module.

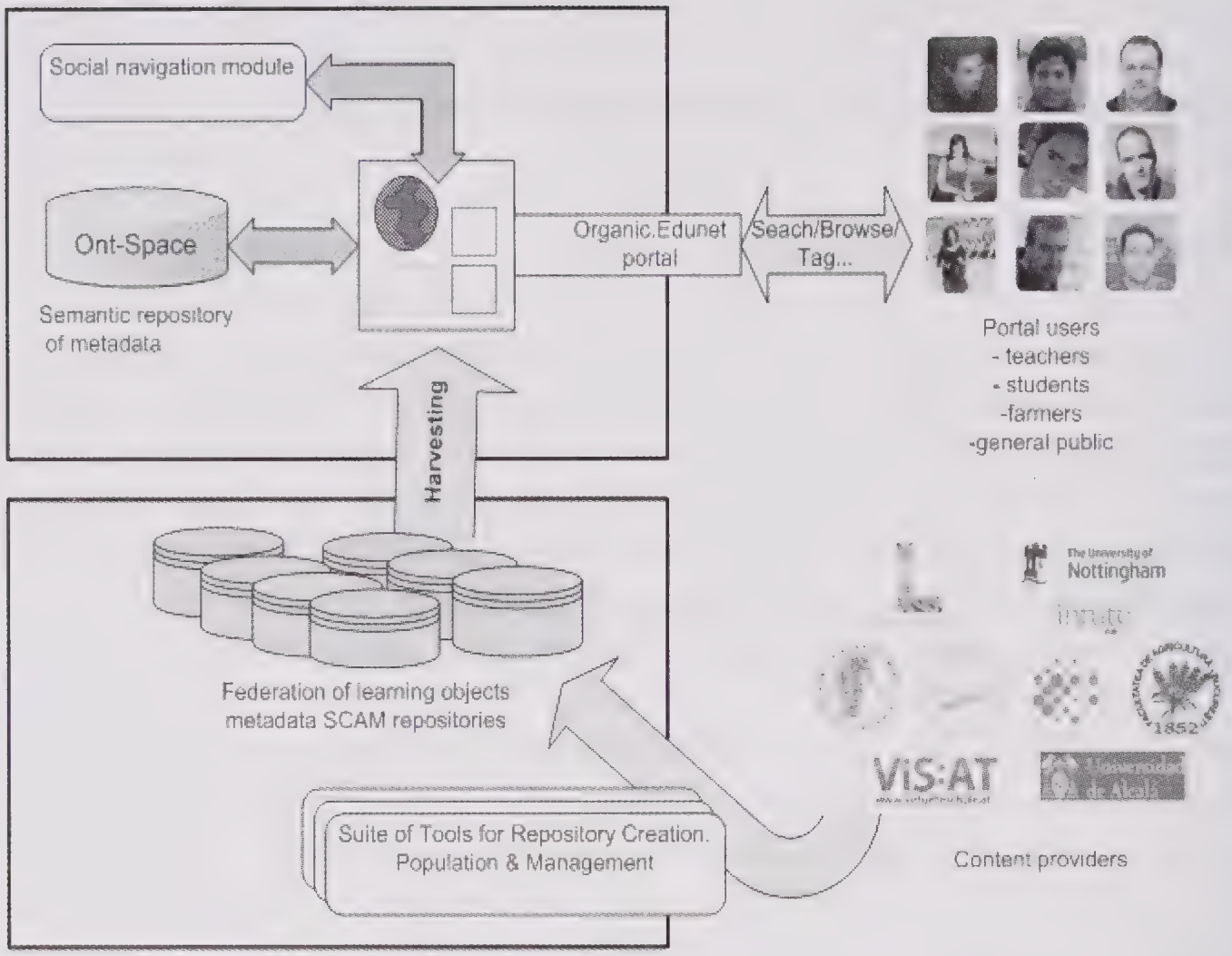


Fig. 1 – Overview of Organic.Edunet technical architecture.

2. 4 Semantic Repository Ont-Space

Ont-space is a software framework for the deployment of semantic Learning Object Metadata Repositories. Ont-space is based on LOMR [3] a semantic learning object metadata repository developed as part of the LUISA project. The semantic repository is driven by ontologies that contain the specification of the learning object schema used such as Dublin Core, IEEE LOM, etc. Ont-space enables specialized components, such as custom query managers and result composers, to benefit from the availability of different, heterogeneous ont-space instances.

2.5 Semantic Navigation Module

The semantic navigation module is responsible for the semantic search capabilities integrated in the architecture. The main objective of the semantic navigation module is to provide an ontology-based semantic interface for performing interactive searching sessions. This is carried out through the navigation of ontology elements (classes, individuals and properties) and their relations. This process is totally ontology-independent, so any valid ontology in OWL format can be navigated, and particularly the OA&AE ontology developed for Organic.Edunet [4]. This type of semantic search provides more accurate and meaningful results, eliminating non-relevant results and avoiding the existence of long lists with repeated elements in the search outcomes.

2.6 Portal Infrastructure

The Organic.Edunet Web portal has been designed to be used by non-expert users (usability has been a priority) and developed over a customized Joomla CMS (Content Management System) installation. The technical partners of the project developed some Joomla add-ons in order to provide customized functions in the Organic.Edunet Web portal architecture.

As the Web application based on Joomla needs to communicate with other architecture elements, there is a need to establish some communication guidelines. The interaction between the Ont-Space module and the Joomla-based Web portal is asynchronous and it takes place through the new components developed and integrated in Joomla. Those components have been implemented in PHP and AJAX, follow the Joomla MVC guidelines. The communication between Ont-Space and Joomla is based on the exchanging of JSON objects that contains the information sent from Ont-Space in an asynchronous way as result of a request from Joomla to a concrete remote servlet in Ont-Space.

2.7 Social Navigation Module

The Web Services API of the Organic.Edunet Social Navigation Module are available only within the scope of the OrganicEdunet's suite of tools. The available methods of this Module's API that are exposed as Web Services are grouped in 5 categories: General Methods, Resource Annotation Methods, Prediction Methods, Recommendation Methods, and Population Methods.

3 Discussion

The Organic.Edunet Tools are fully deployed and are being used by the Organic.Edunet partners. More specifically, the repository tool has been set up and is used by the project partners to upload resources and links, also annotating them with metadata. Presently, a translation of the repository tool is underway to include more languages (i.e. French, Czech, Slovenian, Bulgarian, Turkish, Dutch and Hindi).

The Organic.Edunet Metadata Application Profile is also completed and deployed in the repository tool. Some work is also being carried out for the translation of the AP, so that the repository tool can be set up for different languages, thus helping additional communities connect their material with Organic.Edunet. The same languages that are being added for the repository tool are also translated for the metadata AP.

As far as the Organic.Edunet Portal is concerned, this is currently being used by a great number of users, as the Organic.Edunet Open Days are underway. The participants of the Open Days are also asked to provide their feedback using an online questionnaire. All the results are being evaluated by the responsible Organic.Edunet partner, making the necessary revisions to the portal. Additional languages are also being deployed, building on the existing support offered by the Joomla community. More specifically, all the portal texts

are being translated to open the Organic.Edunet Portal to stakeholders that are interested in connecting their collections to Organic.Edunet Portal.

4 Conclusion

This paper presented the basic tools – technologies that were developed during the Organic.Edunet project. Its main aim was to provide a brief overview of the whole system deployed through Organic.Edunet that can serve as a reference point for institutions trying to open their content to the world, by setting up learning repositories and by making their content available through open access portals. A possible direction to take advantage of the finding of this paper, is to adapt the process to biodiversity, in order to explore how well will this approach fit to this domain.

Acknowledgements

The work presented in this paper has been funded with support by the European Commission, and more specifically the project No ECP-2006-EDU-410012 "Organic. Edunet: A Multilingual Federation of Learning Repositories with Quality Content for the Awareness and Education of European Youth about Organic Agriculture and Agroecology" of the *e*Content*plus* Programme.

References

[1] H. Ebner, N. Manouselis, M. Palmér, F. Enoksson, N. Palavitsinis, K. Kastrantas and A. Naeve, "Learning Object Annotation for Agricultural Learning Repositories", in *Proc. of 9th IEEE International Conference on Advanced Learning Technologies* (ICALT2009), Riga, Latvia, accepted for publication.

[2] S. Ternier, E. Duval, "Interoperability of Repositories: The Simple Query Interface in ARIADNE", *International Journal on E-Learning*, vol. 5(1), pp. 161-166, 2006.

[3] M. A. Sicilia, S. Sanchez, S. Arroyo and S. Martín-Cantero, *Deliverable D4.3: LOMR architectural prototype specification (D4.3). LUISA Learning Content Management System Using Innovative Semantic Web Services Architecture (IST- FP6 - 027149)*. Madrid, Spain: Atos Origin, 2007.

[4] A. Steen, S. Sanchez-Alonso, M. A. Sicilia and G. Lieblein, *ECP-2006-EDU-410012 Organic. Edunet Deliverable D2.2.3, Initial version of Organic Agriculture and Agroecology Domain Model Representation*. Online at: http://www.organic-edunet.eu/organic/files/document/OrganicEdunet _D2.2.3a_final.pdf, 17 November 2008.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 417-418.
ISBN 978-88-8303-295-0. EUT, 2010.

JSTOR Plant Science

Michael Sean Gallagher

Abstract — JSTOR Plant Science is an online environment that brings together content, tools, and people interested in plant science. It provides access to foundational content vital to plant science – plant type specimens, taxonomic structures, scientific literature, and related materials, making them widely accessible to the plant science community as well as to researchers in other fields and to the public. It also provides an easy to use interface with powerful functionality that supports research and teaching, including the ability to measure and record plant specimens, share observations and objects with colleagues and classmates, and investigate global plant biodiversity.

Index Terms — tools, database, resources, type specimens.

JSTOR Plant Science is an online environment that brings together content, tools, and people interested in plant science. It provides access to foundational content vital to plant science – plant type specimens, taxonomic structures, scientific literature, and related materials, making them widely accessible to the plant science community as well as to researchers in other fields and to the public. It also provides an easy to use interface with powerful functionality that supports research and teaching, including the ability to measure and record plant specimens, share observations and objects with colleagues and classmates, and investigate global plant biodiversity.

JSTOR Plant Science strives to be a comprehensive online research tool for aggregating and exploring the world's botanical resources, thereby dramatically improving access for students, scholars, and scientists around the globe. It is useful for those researching, teaching or studying botany, biology, ecology, environmental and conservation studies.

A significant portion of the content available on JSTOR Plant Science has been contributed through an effort known as the Global Plants Initiative (GPI). GPI is an international undertaking by leading herbaria to digitize and make available plant type specimens and other holdings used by botanists and others working in plant science every day. Partners include more than 150 institutions in 52 countries on 5 continents. There are two partner networks in place and contributing today: the African Plants Initiative which focuses on plants from Africa and the Latin American Plants Initiative which contributes plants from Latin America. GPI and has also expanded to Asia. GPI has received funding and guidance from The Andrew W. Mellon Foundation.

M. S. Gallagher, 100 Campus Drive, Suite 100 - 08540 - New Jersey, USA, E-mail: michael.gallagher@jstor.org.

Currently, JSTOR Plant Science has more than 900,000 specimens. When complete, there will be an estimated 2.2 million. Further there are foundational reference works and books such as The Useful Plants of West Tropical Africa, Flowering Plants of South Africa, and illustrations from Curtis's Botanical Magazine. JSTOR Plant Science also includes a significant set of correspondence, including Kew's Directors' Correspondence which included hand-written letters and memorandum from the senior staff of Kew from 1841 to 1928. The JSTOR Plant Science team is developing tools to extract contextual data from this aggregation, which will hopefully enhance its use to the botanical community.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – p. 419.
ISBN 978-88-8303-295-0. EUT, 2010.

ecoBalade: Towards a workflow for Citizen Science Nature Trails

Julie Chabalier, Khaled Talbi, Patrick Peters, Amandine Sahl,
Olivier Coullet, Olivier Assunçao, Olivier Rovellotti

Abstract — In the context of Citizen Science, where potential users may not be familiar with traditional keys and criteria, it seems necessary to provide a simplified interface in order to guide the users through the complexity of natural biodiversity.
Tacit knowledge [1] necessary for citizen science projects; is well known to be difficult to transfer to another person by means of writing it down or verbalizing it. The ecoBalade solves that constraint by putting the user in a problem solving situation in the context of determination, observation, and recording of field data.
In order to be successful the workflow must include three stages; a filter of the potential taxa by an expert naturalist, support thought-out the first identification process and a finally visualization of the field observations by the subject. In a typical ecoBalade scenario, an expert naturalist will survey the potential species beforehand, and will generate a list of potential taxa criteria and pictures. This is then formalized in a XML semantic structure used by a PDA software Pocket eReleve [2] to guide the users thought the identification process.
This concept has been successfully tested on the field in Saint Mandrier with approximately twenty novice users and three PDAs, it has generated in the course of two hours 32 observations [3]. The experience will be generalized to other locations in the months to come.

Index Terms — citizen science, education, biodiversity, XML standard, mobile solutions, identification, survey.

References

[1] M. Polanyi, "The Tacit Dimension", New York, Anchor Books. (108 + xi pp.), 1967.
[2] G. H Kipré and al., "Pocket eReleve, Nouvelle approche de collecte de données sur le terrain" Geomatique expert, vol. 69, pp. 24-27, 2009.
[3] "Prototypage de l'Eco-balade … Et plus si affinités …", http://territoiresenresidences wordpress.com/2010/06/24/.

All authors are with the company Natural Solutions, Marseille, CP 13002.
E-mail: olivier_rovellotti@natural-solutions.eu.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 421-422.
ISBN 978-88-8303-295-0. EUT, 2010.

Electronic data recording tools and identifying species in the field

Alexander Kroupa, Anke Hoffmann, Juan Carlos Monje, Christoph L. Häuser

Abstract — The "European Distributed Institute of Taxonomy" (EDIT) is an initiative of 28 European, North American and Russian institutions to build a network in "Taxonomy for Biodiversity and Ecosystem Research", with the objective to reduce the fragmentation in taxonomy through institutional integration in Europe (www.e-taxonomy.eu). European Commission funding (FP6) for this "Network of Excellence" has started in March 2006, and runs for 5 years. For EDIT Work Package 7 (WP 7) "Applying Taxonomy to Conservation" the aim is to strengthen the input of taxonomic expertise in Europe for biodiversity conservation, by organizing the participation of individual taxonomists and experts in biodiversity inventory and monitoring efforts in conservation areas (www.atbi.eu).

For biodiversity inventories and monitoring, the advantage of using digital field recording tools is to simplify data recording as well as to improve data quality. The use of electronic field tools and software should be promoted to help minimizing error rates, in particular to avoid mistakes from the beginning of the recording chain. Many errors may be avoided by using authority lists, e.g. for countries, habitat-types or taxa that can already be determined in the field. Automated geo-referencing and recording of date and time in standardized formats already in the field will also avoid errors when importing or retyping such data into a database. Relevant software should be usable for tools such as mobile phones with GPS (Global Positioning System) functionality up to water resistant PDAs - Personal Digital Assistant (e.g. Magellan - Mobile Mapper; Trimble – Juno, Nomad).

Examples presented here for more efficient electronic data recording in the field include the application of mobile recording devices with customized forms, which are tested for field work in ATBI+M (All Taxa Biodiversity Inventories + Monitoring; www.atbi.eu) sites operated in the EDIT project. This is a general approach for recording geo-

A. Kroupa is with the Museum für Naturkunde, Invalidenstr. 43, 10115 Berlin. E-mail: alexander.kroupa@mfn-berlin.de.
A. Hoffmann is with the Museum für Naturkunde, Invalidenstr. 43, 10115 Berlin. E-mail: anke.hoffmann@ mfn-berlin.de.
C.J. Monje is with the Staatliches Museum für Naturkunde, Rosenstein 1, 70191 Stuttgart. E-mail: carlos.monje@smns-bw.de.
C.L. Häuser is with the Museum für Naturkunde, Invalidenstr. 43, 10115 Berlin. E-mail: christoph.haeuser@mfn-berlin.de.

referenced, individual species data using customized forms for ESRI ArcPad applications. Species names can be selected from a taxonomic authority list provided in a file in dBASE-format. Such files can be easily created, modified, and exchanged to allow individual researchers to use regional or otherwise customized species lists. Fields and field formats correspond to ABCD standards so that exports of recorded locality, event, and species data can be directly integrated into a central database and applications for individual ATBI+M websites (e.g. www.atbi.eu/mercantour-marittime/ or www.atbi.eu/gemer/). The authority species lists may be customized for a geographic area (e.g., a nature reserve) and/or a group of taxa (e.g., larger birds). This allows each expert to choose the species list needed for his/her research. Problems remain with observation records which cannot be reliably determined in the field. Therefore identification help should be made available on the PDA at least for difficult taxa.

Index Terms — biodiversity, digital data capture, fieldwork, inventory, species authority lists.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 423-428.
ISBN 978-88-8303-295-0. EUT, 2010.

Cost Assessment of the Field Measurement of Biodiversity: a Farm-scale Case Study

Stefano Targetti, Davide Viaggi, David Cuming

Abstract — Attention to the effects of agriculture on biodiversity is currently increasing. Yet the measurement of biodiversity is both time-consuming and costly. Considering the limited budgets available for biodiversity conservation, it is timely to focus on the cost analysis of biodiversity indicators in order to ensure the optimization of the scarce funds available. We present the cost analysis of operational data from the fieldwork efforts undertaken in the measurement of biodiversity indicators at farm-scale. Methodological issues are discussed.

Index Terms — biodiversity measurement, earthworm indicator, spider indicator, research costs.

1 Introduction

The growing societal demand for environmental services provided by agriculture focuses attention on the implementation of sound agro-environmental schemes based on reliable information with respect to the effects of different agricultural practices on biodiversity [1]. The existing gap between need and availability of funds for biodiversity monitoring and assessment highlights the importance of the optimization of resources [2]. The cost analysis of biodiversity measurement, in particular if undertaken by way of a cost-effectiveness analysis, can ensure the optimization of scarce available funds and the selection of the most efficient indicators of biodiversity [3], [4]. Nevertheless, the cost-effectiveness of biodiversity measurement is a practically unstudied issue [5] and only a few examples exist [6] which propose a methodological approach to its analysis.

The assessment of the costs of measuring biodiversity at farm-scale is one of the specific tasks of the BioBio project (Indicators for biodiversity in organic and low-input farming systems -UE-FP7- http://www.biobio-indicator.wur.nl). In this paper we propose a methodology for the cost assessment of biodiversity

The authors are with the Department of Agricultural Economics and Engineering of the University of Bologna, Italy. E-mail: stefano.targetti@unibo.it, davide.viaggi@unibo.it, david_cuming@hotmail.com.

measurement, and discuss its practical application to the spider and earthworm indicators measured through the BioBio project protocol.

2 Material and methods

The BioBio project involves 12 case studies (CS) throughout Europe concerning organic, or low input, and conventional agricultural systems. Here we present the preliminary results from the French CS (Midi- Pyrénées Region). The cost assessment is related to the field measurement of the spider (SP) and the earthworm (EW) biodiversity indicators carried out in Spring 2010. Data covers 4 arable farms (Tab. 1), where wheat and sunflower are the main crops. Distance from the research centre (driving time in minutes) was similar for each farm (about 1 hour). Survey stratification was performed through habitat mapping in order to cover the different habitat conditions of the surveyed farm sites (see [7] for further details). In the aggregate, 537 samples (345 for SP and 192 for EW) were gathered.

Farm	Area (ha)	Type	Distance from research centre (minutes)	Number of samples
A	23	organic	53	113
B	19	organic	57	88
C	27	conventional	60	111
D	146	conventional	68	225

Tab. 1 – Main features of the 4 farms studied and number of samples (SP + EW) gathered during the spring fieldwork.

Spider sampling was carried out with the aid of a modified vacuum shredder (Stihl SH 86-D), and 5 suction samples were taken in each plot (each suction had a suction area of approximately 0.1 m² and lasted 30 seconds). The samples were stored separately in a cool-box and transferred to a laboratory. Spiders were sorted out in the laboratory and placed in vials with 70% alcohol [8], [9]. Three survey sessions were scheduled in the project protocol. Here we present data from the first session. The sampling team was composed of 3-4 persons.

Earthworm sampling was carried out by way of two methods: 1) stirring up an allyl-isothiocyanate and ethanol solution into metal frames (30 cm X 30 cm) which were placed in the ground, and collecting the earthworms that came upward during the first 10 minutes; 2) extracting the soil core (20 cm depth) from the sampling site and hand-sorting the earthworms on a plastic sheet. Samples were placed in cold containers with oxygenated water and transferred to refrigerators in the laboratory [10], [11]. The sampling team was composed of 5 persons.

The cost assessment methodology was organised in such a way as to allow for an analytical assessment of actual costs, as well as the subsequent simulation of costs with standardised costs. For this reason, both physical units

of resources used and related prices were collected on a regular basis. Data collection was performed through the collection of records related to staff time, distance and transport time, consumables and equipment. Time spent (and costs) for fieldwork organisation and preparation and taxonomy identification is not included. Field staff filled-in a weekly cost-form which was entered into a relational data-base. Data collection was organised in order to retrace the costs related to each single farm and each single indicator. Each record contained: date, farm site, staff qualification, time spent per field-worker and was linked to different tables indicating the salary band of the staff, the distance of the farm site from the research centre, transport time, equipment and consumable costs, and the type of work (fieldwork, laboratory, etc.). The cost of the indicator measurement was composed of three resource categories: 1) equipment and consumables, 2) labour time investment (fieldwork, laboratory-work and transport), 3) worker categories (permanent, temporary).

Equipment and consumables included all the materials used during the fieldwork as well as the field lunches for the staff. The cost of the vacuum shredder was calculated as: cost per suction = cost of the vacuum new / number of suctions over its lifetime. This was approximated to 0.038€ per suction. The gross salary of the staff was approximated to 36€ per hour for permanent workers and 13.8€ per hour for temporary workers. Vehicle costs were charged at 0.32€ per km and included fuel, car insurance and vehicle depreciation. All the costs are related to 2010.

3 Results

The composition of the costs for the field measurement of the two biodiversity indicators in the four farms studied are presented in Tab. 2. The cost per sample of the earthworm indicator was 3.5 times higher than the spider indicator. EW costs per hectare were only 2 times higher than SP because of the lower number of samples gathered for the EW indicator. Although the spider indicator required a higher permanent work effort (1 hour of permanent work for every 2.6 hours of temporary work for SP vs. 1 hour of permanent work for every 4 hours of temporary work for EW), the labour load was higher for the EW indicator (the labour cost was 83% of total cost for EW vs. 57% for SP). The portion of the other costs were always lower (max. 10% of total costs), except for lab work and preparation of samples which constituted an important component of costs for the spider indicator (23% of total costs).

The cost of transportation (vehicle, highway tolls and work time for transfer of fieldworkers from the research centre) was a consistent portion of costs for the measurement of biodiversity (Tab. 3). This cost was about 30% of total costs. Accordingly, the cost of the measurement of the indicators was strongly tied to the organisation of the fieldwork (number of sessions, distance of farms from research centre, etc.). The portion of transportation + transfer of fieldworkers with respect to the total costs was higher for SP than for EW (34% vs. 28%) because the research unit was equipped with only one vacuum tool. As a result only one sampling team could be organised for fieldwork each day. The EW measurement was more flexible as several sampling teams per day could be

arranged. Thus, the differences in costs between the two indicators were more evident when considering the effective costs of fieldwork (resources spent in field measurement after transport costs): 13.3€ ha^{-1} for SP vs. 28.1€ ha^{-1} for EW (ratio 1:2.1).

Biodiversity indicator	Spiders	Earthworms
Cost per sample	12,6	44,3
Cost per ha	20,2	39,6
Labour	618	1756
Permanent / Temporary ratio	1: 2,6	1: 4
Consumables and equipment	110	187
Labwork	253	35
Vehicle and tolls	105	148
Sum of costs	**1086**	**2126**

Tab. 2 – Composition of costs (mean values per farm) for the field measurement of biodiversity indicators (values in €) and permanent vs. temporary work effort ratio.

Biodiversity indicator	Transport costs (vehicle + displacement of fieldworkers, €)	Percentage of total costs (%)	Effective cost per sample (€)	Effective cost per ha (€)
Spiders	369	34	8,3	13,3
Earthworms	618	28	31,4	28,1

Tab. 3 – Analysis of costs of the field measurement of biodiversity. Share of transportation and transfer of fieldworkers with respect to total costs (mean values per farm) and effective costs of fieldwork (effective costs are: total costs – transport and transfer of fieldworker costs).

The comparison of effective costs of biodiversity measurement between organic and conventional farms pointed out a consistent higher effort of field sampling for the organic farms (Tab. 4). Even if the mean number of samples was higher in the conventional farms (84 vs. 50), sampling effort in organic farms was 1,5 times higher concerning the cost per hectare and 2,4 times higher considering the days $person^{-1}$ ha^{-1}. This is probably related to a higher variability of habitats for the organic farms which required a more intense sampling than the conventional farms.

	Samples	Effective cost ha^{-1}	Effective days per person ha^{-1}
Organic	50	41,6	0,62
Conventional	84	27,5	0,26

Tab. 4 – Comparison of costs of field sampling in organic and conventional farms. Number of samples, effective cost ha^{-1} and effective days person^{-1} ha^{-1} of effort required (mean values of Sp + EW sampling per farm, effective cost is: total cost – transport and transfer of fieldworker costs).

4 Conclusion

The first important result concerns the relevance of costs that were in the thousands of Euros per farm.

The *ex-post* assessment of costs of the field measurement of biodiversity is of significant importance both for the organisation of the sampling sessions as well as for the cost-effectiveness analysis. The cost assessment could be a valid tool for the optimisation of the use of available resources. This evidence is of great importance considering the gap between the need and the availability of funds for biodiversity. It is our opinion that the increased availability of cost data could be of great assistance in the advancement of the effectiveness of biodiversity assessments.

The share of transportation costs (vehicle and transfer time of staff) suggests that a careful organisation of fieldwork should be considered essential for the optimisation of available resources.

Our preliminary analysis clearly identified lower costs, coupled with a higher number of samples (thanks to the vacuum tool), for the spider indicator. However, this information is incomplete without an assessment of the effectiveness of the measurement. Moreover, the cost of SP will be much higher considering the other two survey sessions which are scheduled in the BioBio project protocol.

Acknowledgement

This work was supported by a grant from EU-FP7, BioBio - Indicators for biodiversity in organic and low-input farming systems. The authors wish to thank J.P. Sarthou, J.P. Choisis, C. Pelosi and S. Ledoux for their cost data gathering.

References

[1] OECD, "OECD Expert Meeting on Agri-biodiversity Indicators", Zurich, 5-8 November 2001, http://www.oecd.org/dataoecd/16/56/40339943.pdf, 2010.

[2] T. B. Gardner, J. Barlow, I. S. Araujo, T. C. Avila-Pires, A.B. Bonaldo, J. E. Costa, M. C. Esposito, L. V. Ferreira, J. Hawes, M. I. M. Hernandez, M. S. Hoogmoed, R. N. Leite, N. F. Lo-Man-Hung, J. R. Malcolm, M. B. Martinus, L. A. M. Mestre, R. Miranda-Santos, W. L. Overal, L. Parry, S. L. Peters, M. A. Ribeiro-Junior, M. N. F. Da Silva, C. Da Silva Motta and C. A. Peres, "The Cost-Effectiveness of Biodiversity Surveys in Tropical Forests", *Ecology Letters,* vol. 11, pp. 139-150, 2008.

[3] P. J. Ferraro and S. K. Pattanayak, "Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments" *Plos Biology,* vol. 4, pp. 482-488, April 2006, http://

www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.0040105, 2010.
[4] B. S. Halpern, C. R. Pyke, H. E. Fox, J. C. Haney, M. A. Schlaepfer and P. Zaradic, "Gaps and Mismatches Between Global Conservation Priorities and Spending" *Conservation Biology,* vol. 134, pp. 96-105, 2007.
[5] A. Juutinen and M. Mönkkönen, "Testing Alternative Indicators for Biodiversity Conservation in Old-Growth Boreal Forests: Ecology and Economics", *Ecological Economics* vol. 50, pp. 35-48, 2004.
[6] A. Qi, J. N. Perry, J. D. Pidgeon, L. A. Haylock and D. R. Brooks, "Cost-efficacy in Measuring Farmland Biodiversity – Lessons from the Farm Scale Evaluations of Genetically Modified Herbicide-tolerant Crops" *Annals of Applied Biology,* vol. 152, pp. 93-101, 2008.
[7] R. Jongman and R. G. H. Bunce, *Farmland Features in the European Union. A Description and Pilot Inventory of their Distribution*, Alterra report 1936, ALTERRA, Wageningen UR, 2009.
[8] M. H. Schmidt-Entling and J. Dobeli, "Sown wildflower areas to enhance spiders in arable fields", *Agriculture Ecosystems & Environment,* vol. 133, pp.19-22, 2009.
[9] M. H. Schmidt and D. T. Tscharntke, "The Role of Perennial Habitats for Central European Farmland Spiders" *Agriculture Ecosystems & Environment,* vol. 105, pp. 235-242, 2005.
[10] E. R. Zaborski, "Allyl isothiocyanate: an alternative chemical expellant for sampling earthworms", *Applied Soil Ecology,* vol. 22, pp. 87-95, 2003.
[11] C. Pelosi, M. Bertrand, Y. Capowiez, H. Boizard and J. Roger-Estrade, "Earthworm collection from agricultural fields: Comparisons of selected expellants in presence/absence of hand-sorting", *European Journal of Soil Biology,* vol. 45, pp. 176-183, 2009.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 429-435.
ISBN 978-88-8303-295-0. EUT, 2010.

Markets for biodiversity information products: real or imaginary?

Bill Hominick, Peter Schalk

Abstract — In the past decade, a large range of biodiversity information products and services has become available. Some were developed with EC subsidies, some with national funds, and others as initiatives of universities, research institutes or private persons. Few, if any, were developed in the 'commercial world' based on a business plan. Some sponsors and funders ask for exploitation and sustainability plans after the development phase or at the end of the project. The success of these depends on whether the products meet real demands and serve actual markets. Market potential is often misjudged or overestimated and many products are developed without a prior needs or market analysis. This negatively affects the sustainability for biodiversity information services. In this paper we review the marketing of some biodiversity information products in a commercial environment, to assess demands and size of the markets.

Index Terms — biodiversity information products, user needs, markets, return on investment, sustainability.

1 ETI Information Services: a specialized sales Company

ETI Information Services Ltd (ETIIS) was a subsidiary of ETI BioInformatics, a not-for-profit organisation initiated by the Netherlands' Government and UNESCO. Its aim is to make authoritative biodiversity information broadly accessible and usable by using information technology. Initially Springer Verlag marketed and distributed ETI information products, but recognising the unique market for e-media and specific requirements to reach the audience, ETI developed its own marketing and distribution subsidiary. ETI Information Services Ltd was established in 2001 and relied on the catalogue on the website of ETI. The company launched its own website to support sales on 1 April, 2004. The company ceased trading on 31 March, 2009. ETI products are currently marketed by Margraff, Germany. The purpose of this paper is to share the commercial experience gained in marketing electronic biodiversity

B Hominick was Executive Director of ETI Information Services Ltd, UK (2001-2009). E-mail: bill@etiis.org.
P Schalk is managing Director of ETI BioInformatics, University of Amsterdam, Netherlands. E-mail: peter@eti.uva.nl.

information products, to discuss the size of this niche market, and the marketing requirements to reach it. The analysis is based on commercial sales of selected multimedia biodiversity titles to 1705 customers from over 50 countries from 2001 to 2008 in a period of 77 months. During that time 9970 items were sold. Sales were direct to customers online, by mail, e-mail or phone orders, or through re-sellers reaching particular markets. All resellers sold multimedia products as a minor part of their main business, which was usually selling books. ETIIS relied exclusively on income from the niche market for electronic resources related to biodiversity.

The income generated by sales is compared with product development costs to get an impression of the sustainability of such products.

2 Users of Biodiversity Information

ETIIS focused on three categories of users of biodiversity e-products: Teachers and students in schools, colleges and universities (Formal Education); outdoor enthusiasts such as walkers, wildlife observers and workers in Nature reserves (Popular); and, professional specialists, researchers, regulatory officials and dedicated amateurs (Science). Each of these markets has different sizes and requirements and needs to be considered individually. The Formal Education market is not considered in this communication.

3 Popular Market

The general population is a large and important market for biodiversity information. A number of ETI wildlife field guides should be of interest to non-specialists, i.e. individuals who have an interest in some aspect of nature, and are committed to learn more. This is a highly competitive market with numerous printed products available, so the e-products must be priced competitively. The following titles, showing the price range (including VAT) charged by different distributors, fit these criteria:

- Birds of Europe (£14.95 - £25.95)
- Interactive Guide to Butterflies of Europe (£12.95 - £25.95)
- Interactive Guide to Mushrooms and other fungi (£12.95 - £25.95)
- Interactive Flora of the British Isles (£29.95 - £39.95)
- Five Kingdoms – a multimedia guide to life on earth (£14.95 - £25.95)

ETIIS used several resellers specialized in particular markets to sell some or all of these titles. Their sales should indicate the interest of their customers in biodiversity information in an e-format. Tab. 1 lists the main resellers and sales data, and is followed by some observation on sales and markets.

DISTRI-BUTOR	MARKETING	BIRDS 2.0	IFBI	FIVE KD	BUTTR FLIES	MUSH RM
		1992-2006	May 04	May 02	Jan 02	Dec 02
Alana Ecology	UK :Web, catalogue, adverts, attend fairs	33	261	10	45	53
DreamDirect	UK: Catalogues, Web	1950 Jan04-06	NS	NS	60 (Jan-Feb 04)	NS
NHBS	International: Web, catalogue, trade fairs, meetings	3	167	30	63	18
RHS	S England: Retail shop Wisley Gardens	14	58	27	10	26
Summerfield	UK: Web, catalogues, attend meetings	NS	502	NS	NS	14
ETI IS	Web, adverts, fliers, meetings	99	486	145	239	76
TOTALS		**2095**	**1474**	**212**	**417**	**187**

Tab. 1 – Number of e-guides sold by ETIIS and resellers with information on publishing date and marketing methods. NS = Not Stocked.

Observations regarding Tab. 1:

1) The best-selling popular title by far is Birds of Europe. *DreamDirect* markets carefully-selected products to a mass market by distributing up to 36 million catalogues per year. Remarkably, Birds of Europe is one of the first titles produced by ETI and over ten thousand (in various updates) have been sold since 1992.

2) In contrast to DreamDirect's mass marketing approach, *Alana Ecology* is specialized, and targets outdoor enthusiasts. It features books and CD-ROMs as a subset of its products, which are mainly equipment. For their customers the Interactive Flora of the British Isles (IFBI) is a favourite; they had less interest in birds, butterflies and mushrooms, field guides that could be used by this group.

3) *NHBS* supplies shops, institutions and private customers. It lists all ETI titles amongst the book categories; customers have to seek them out. Though they have numerous books and field guides on birds, they sold only 3 copies of Birds of Europe. Are paper field guides the preferred medium for these customers? IFBI and Butterflies are their best lines for ETI products, promoted strongly in their catalogue and website. Sales indicate that most of their e-products achieve very small sales.

4) The *RHS* has a large, customer base of gardeners and horticulturalists and is known for its large selection of plant titles. The shop has a small number of e- products, and does not market them actively. IFBI was their best-selling ETI title. The other sales are likely to be impulse buys.

5) *Summerfield Books* targets the plant sciences market, and e-products are uncommon in their catalogue. However, they actively market titles and have achieved strong sales of IFBI. The BSBI strongly supported the IFBI. Apparently, an audience can be persuaded to purchase these e-products if they are aimed specifically at them and their field of interest.

6) Sales by *ETIIS* itself demonstrate the importance of marketing in boosting sales. In this selection, IFBI is the best-selling popular product. It was heavily promoted and received strong reviews in popular and scientific journals; it is now recognized as one of the standard (e)flora for the UK. Advertisements were taken in relevant publications (eg Plant Talk) and it was promoted via Google Adwords. The Interactive Guide to Butterflies of Europe is the other best-seller of the popular titles. It was promoted mainly by a repeated advertisement in *Butterfly*, the magazine of Butterfly Conservation (readership 17,000). However, such promotion is expensive and the cost of promotion exceeded the income generated by increased sales!

Sometimes it is implied that e-products are preferred over printed versions. This is not the case. For example, compare the book *Flora of the British Isles* First Edition: Sales 7400 (1991-1997) and Second Edition: Sales 7350 (1997-2004) with the Interactive Flora of the British Isles DVD-ROM: Sales 1474 (2004-2008). Similarly, the book *Flora of the Netherlands* print run was 18,000 while Heukel's Interactieve Flora CD-ROM sold over 5,000 copies in four years. E-products have a higher access barrier than books, and the interest of retail shops in stocking e-titles is limited as the products are considered too specialized and with limited sales potential. This prejudice towards the print medium must be overcome if sales for biodiversity e-information are to increase.

We conclude that there are many potential users of biodiversity information in the general population, and some are willing to pay for and use it in an e-format. Amongst the 612 biodiversity information e-product ETIIS sold, birds were by far the most popular subject. Price, within a limit of £40, does not appear to be an issue for the customer. However, the single most important fact is that while there are many potential users of biodiversity e-information, they need to be made aware of the existence of the product, and then be persuaded to purchase it. Most are not actively seeking to purchase multimedia biodiversity information. Hence marketing is critical. Of course, if the information was free, the situation could be different. ETI's website 'soortenbank.nl' freely offers detailed information, identification keys and distribution maps on almost 7,000 species in the Netherlands. It attracts well over 3,000 unique users daily, a number that is still growing.

4 Science Market

The science market is fragmented and specialized, with small sales to be expected for the vast majority of similarly specific e-products. The largest category of products is from ETI's World Biodiversity Database series: 80+ e-publications. These e-products are taxonomic monographs aimed at specialists and by their nature will have a limited market. It is difficult and costly to reach a small audience. Sales do not appear to be strongly price-sensitive. Like books, sales patterns show that most sales are achieved in the first few years for a title, and then enter a steady state of a few sales per year. Reducing prices can have a temporary effect on lagging sales. Special bulk sales, at a discount, can have a large affect on sales. Specialist training courses are obvious targets, but usually rely on the author's support or knowledge of training efforts. Authors

can be extremely helpful and work hard to promote their publication. They supply mailing and e-mailing lists for contacts, contact their colleagues and promote their titles at relevant meetings. We collated information for the best-selling scientific titles in 2002-2008 listed in Tab. 2 together with sales, prices, publication dates and months available.

Product	Publishe don	Months available	Price ex VAT (GBP)	Units sold
World Seaweed Resources (WSR)	01/04/06	24	21.23	243[1]
WSR with Cult. Farming of Marine Plants		24	34.00	97
Yeasts of the World	07/05/02	72	59.95	140[2]
Arthropods Econ. Importance - Diaspididae	13/12/02	64	77.95	134[3]
Otoliths of N Sea Fish	01/11/01	78	59.95	88
Zooplankton South Atlantic Ocean	01/08/05	33	42.51	79
Chironomid Larvae	18/09/03	56	107.95/79.95	62
Chironomid Pupae	14/05/03	60	107.95/79.95	59
Crabs of Japan	28/07/04	46	77.95/42.51	57
Harmful Marine Dinoflagellates	05 08 02	69	29 95	56
Arthropods Econ. Importance - Agromyzidae	28/07/04	46	77.95	56
Freshwater Oligochaetes	24/02/03	61	107.95/79.95	53
Turtles of the World	2000	78	59 95 36.95	52
Zooplankton/Micronekton N. Sea	01/04/03	60	79.95/42.51	51

Tab. 2 – Best-selling science titles of ETIIS.

Observations regarding Tab. 2:

- Arthropod titles were promoted by e-mails to addresses provided by the author and to e-mail lists created by data mining. Specialist bookshops and fliers at relevant meetings were also used to promote the titles.

- Chironomids and Oligochaetes. A remarkable number of sales were achieved for these three products. While these are very specialized groups, they have great importance for environmental monitoring of water quality. This market was reached by building up e-mail addresses of water quality specialists. Titles of practical importance are less price-sensitive as they are frequently “must have” rather than “would like to have” products.

- Although World Seaweed Resources is a fairly recent publication, it is the best-selling science title. It was priced low deliberately to encourage sales, and this was possible because a company provided sponsorship for its production. While the low price undoubtedly helped sales, the author has been very important in promoting it. In addition the product was bundled with Cultivation and Farming of Marine Plants at reduced price, which had a remarkable effect on sales of Cultivation and Farming of Marine Plants, one of the poorest-selling scientific titles, with only 14 sold separately over a period of 77 months compared to the 97 sold as part of the special package.

Scientific e-titles, by their nature, cover very specialized groups and have a small, fragmented market. Even best-selling titles rarely achieve sales beyond 100. In comparison: scientific books usually have print runs of 500-800. For

the foreseeable future, books will outsell e-products even for the same title: books have the clear advantage in portability, comfort of handling, familiarity, shelf life and identifiable prestige on the shelf. The trend in scientific publishing is towards smaller print runs, pitched to a known market and then straight to Publish on Demand. E-products are ideal for this trend, as they can be produced easily at little cost in small numbers, and they can also be updated easily.

After the first few years of sales, there will be a long period when sales continue at a trickle rate. Significantly reducing prices can promote the level of these sales, while linking to e-mail marketing can enhance the effect. However, the costs of producing a database for e-mail marketing are substantial, and are not met by the small numbers of sales. Hence, even if funding is available, one should ask, "Who wants the information?" Always ask, "Is this a need-to-have or a would-like-to-have, product?" All biodiversity information is not equal. It could be argued that it is nice to know about the butterflies or birds in a region, but it is essential to know about pests, crop protection, insect vectors of disease, etc. So, if investment is required, some priorities will also be required.

5 Product devevelopment costs

The selected biodiversity information products discussed here use the *Linnaeus II taxonomic information management system* as a vehicle to compile, organize, share, and e-publish (on disk, web, mobile) biodiversity data. Linnaeus II was developed in the 90-ties (initially with subsidies from UNESCO and the Dutch government) as a freely available data management tool for scientists to create 'e-monographs'. The website: *http://www.eti.uva.nl/products/linnaeus.php* provides further information. This software is updated in 3-4 year cycles. Current version is 2.65 (2008), while a wholly new (web-based) version (Linnaeus NG) is expected in 2011. When calculating the development costs for e-products as described here, the generic costs of (maintaining) the information management system must be taken into account. ETI estimates the cost of an update cycle of Linnaeus to approximately 50,000 Euros. The cost for technical user support (helpdesk) amounts to 15,000 euro per annum. ETI distributes the additional costs equally over all e-products amounting to 5,000 Euros per e-product.

Tab. 3 gives an indication of the product development costs of the popular-educational and scientific e-products discussed. Popular-educational products, created by ETI staff in collaboration with external specialists and contributors, are more expensive to develop than the scientific titles which are often built and delivered by the authors to ETI in an almost publishing-ready state. However, when the author's research costs and institutional overheads are taken into account especially the science products would be extremely expensive.

Title	Costs*	Revenues	Results	
Birds of Europe (1)	120,000	302,000	+ 182,000	
HIFN+IFBI (2)	191,000	97,000	- 93,000	
Butterflies (3)	43,000	51,000	+ 8,000	
Mushrooms (3)	91,000	51,000	- 40,000	
Scientific Titles (5)	20,000	7,000	- 13,000	(av. per title)

Tab. 3 – Overview of developments costs for several products compared to revenues from sales. Notes: (*) includes 5,000 Euros 'system maintenance costs' see text. (1) Birds of Europe was produced in the English, Dutch, German and Italian language. (2) The Interactive Flora of the British Isles built upon and extended an earlier developed Interactive Flora of the Netherlands. (3) The Butterflies and Mushrooms e-guides were produced in the English and Dutch language. The multi-language approach increased the market potential for the content. (5) The development costs for purely scientific titles is based upon the average of the e-products listed in Tab. 2; similar for the revenues.

ETI has been relatively successful in marketing biodiversity information e-products. Still, the development costs of many less popular e-products exceed the income generated by sales as demonstrated by Tab. 3. Sustainability (i.e. updating the products, keeping them available) is therefore an issue. Academic developers not always properly calculate full development costs (i.e. all hours, all overheads) when considering product exploitation and sustainability issues. A way to address this is: cooperation, increased efficiency by standardization of data (exchange) formats, shared software (developments), joint marketing approaches, in combination with serving the markets needs. The *KeyToNature* project (*www.keytonature.eu*) demonstrated that a collaborative approach is possible, as for computer based species identification products.

Acknowledgement

The authors wish to thank Dr Christian Kittl, Prof Pier Luigi Nimis, Mr Nicola Dorigo, Dr Wim Backhuys, Mr Paul van Bruggen, Dr George Tippet for stimulating discussions.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 437-443.
ISBN 978-88-8303-295-0. EUT, 2010.

A Basic Business Model for Commercial Application of Identification Tools

Christian Kittl, Peter Schalk, Nicola Dorigo Salamon, Stefano Martellos

Abstract — Within the three-year EU project *KeyToNature* various identification tools and applications in formal education for teaching biodiversity have been researched and developed. Building on the competencies of the involved partner organisations and the expertise gained in this domain, the paper outlines a business model which aims at commercially exploiting the project results on a broader scale by describing the value proposition, products & services, value architecture, revenue model and the intended market.

Index Terms — business model, identification tools, exploitation, sustainability, EU project.

1 Introduction

From September 2007 to the end of 2010 the *KeyToNature* project mobilises 14 partners from 11 EU countries in the *e*Content*plus* Programme, with a total budget of 4.8 Million Euros. The main objectives of *KeyToNature* are to: 1) increase access and simplify use of e-learning tools for identifying biodiversity, 2) improve interoperability among existing databases for the creation of identification tools, 3) optimise educational efficiency and increase quality of educational contents, 4) add value to existing identification tools by providing multilingual access, and 5) suggest best practices against barriers that prevent the use, production, exposure, discovery and acquisition of the digital contents required for designing the identification tools [1].

Software packages developed in recent decades, which enable the rapid and easy creation of interactive identification tools, are the driving forces behind

C. Kittl is with evolaris next level GmbH, Hugo-Wolf-Gasse 8, A-8010 Graz. E-mail: christian.kittl@evolaris.net.
P. Schalk is with ETI BioInformatics, Mauritskade 61, 1092 AD Amsterdam. E-mail: pschalk@eti.uva.nl.
N. Dorigo is with T&B e Associati srl, c/o AREA SCIENCE Park, Padriciano 99, Building H I-34012, Trieste. E-Mail: dorigo@tbassociati.it.
S. Martellos is with Department of Life Sciences, University of Trieste, I-34127, Trieste. E-mail: martelst@units.it.

the switch from traditional paper-based keys to multimedia and online versions with their many advantages. These interactive identification tools are not only important for the educational sector, but can also be used to solve identification problems in many industrial application areas. The basic assumption of the *KeyToNature* project, namely that these tools should not only be usable by a few experts, but made applicable for pupils and students by reengineering them to fit their needs and wants, also applies when aiming to reach a broad audience of potential customers in industry: usability, support of multiple languages, aesthetic appeal, the possibility to easily enhance or change parts of the derived keys by adding user generated content, etc., are all factors that are equally important, irrespective of whether the tools are applied in the classroom or in professional environments.

In order to exploit the knowledge gained in the project in the best possible way and thus being able to keep the developed services and tools up to date and usable, a sustainable business model is needed. This will ensure that solutions for the educational field, which are already now used by a large number of *KeyToNature* associated members like schools and universities all over Europe, can be provided even after Community funding ends. This could also help interested project partners to generate returns on their investments in the project, as about half of the budget was financed by their own resources.

In order to analyse the potential of a business, traditionally two different approaches were used: 1) the resource-based view, which focuses on the core competencies and the unique access to resources a company has in order to build competitive advantage (for an overview of the most important works see [2]), and 2) the market-based view, which emphasises the industry with its competitors, customer segments and regulations in which the company has to be successfully positioned [3]. Both approaches are still important when developing a new business, although the main innovation often lies in the business model, especially when modern ICT (Information and Communication Technologies) plays a crucial role in the business [4].

Chapter 2 thus first describes the products developed within the *KeyToNature* project in order to better understand the resources which can be used for exploitation when designing the business. The following chapter then introduces a hypothetical business model and briefly outlines the market strategy.

2 Identification Tools and Results of *KeyToNature*

The identification of organisms is fundamental in biology. An accurate identification provides a correct name through which detailed information such as taxonomic descriptions, ecological relations, economic values, conservation status, legislation status and genetic code can be unlocked [5].

Identification is done by observing certain characters of an organism and subsequently using the respective character states in a structured way by applying a key to retrieve the correct name. This approach can not only be applied for identifying organisms in the field of biology, but basically applies to any identification problem. The tools generated and the expertise gained within *KeyToNature* can thus in principle be generalised to other fields of application.

The following figure gives an overview of the main elements of the *KeyToNature* system architecture:

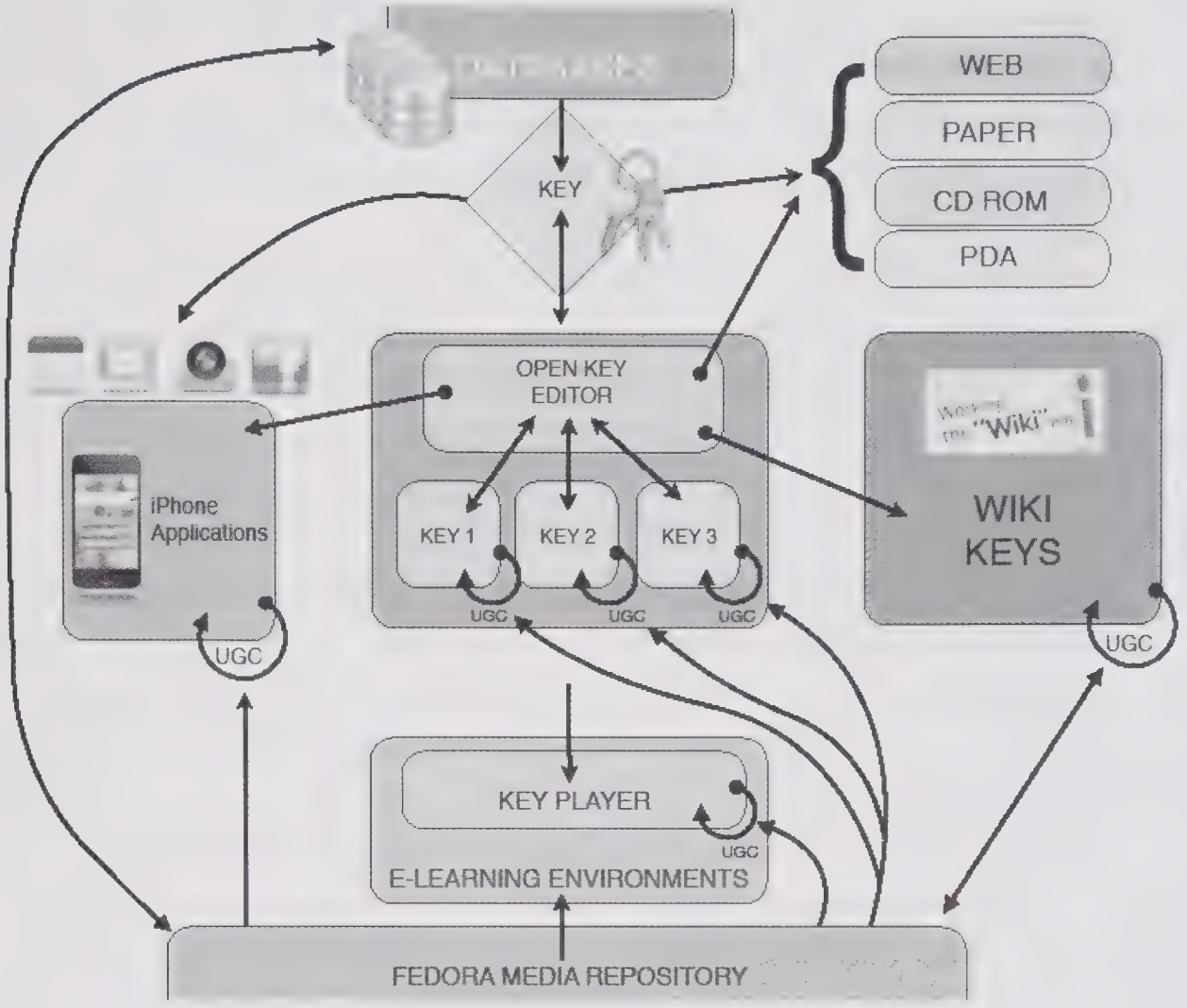


Fig. 1 – The *KeyToNature* system architecture [6].

The heart of the system architecture is formed by the keys. The whole system is primarily aimed to be used online, but the keys can also be made available offline, e.g. for use abroad in the field with mobile phones, where Internet connection fees are expensive or when no Internet connection is available at all. They can thus be web-based, paper-based (i.e. printed out), provided on CD-ROM or on PDAs/mobile devices.

The keys are normally based on data stored in relational databases and software packages like FRIDA and LINNAEUS which dynamically generate the actual sequence of questions step per step. Within the *KeyToNature* consortium the following two tools are being developed by project partners:

FRIDA (FRiendly IDentificAtion) is an original and flexible program developed by S. Martellos and patented (2002) by the University of Trieste. FRIDA is based on a relational database and can automatically generate both interactive identification tools accessible online, and traditional, dichotomous paper-printed identification keys.

LINNAEUS II is developed and sold as a product by ETI. There are three 'modules' of Linnaeus II: the 'Builder' to manage data and to create an information system, the 'Runtime' engine to publish completed information systems on CD-ROM/DVD-ROM, and the 'Web Publisher' to publish a completed project as a Web site.

Besides the primary data (information that generates the actual key) these software packages make use of secondary data like pictures, drawings, sound

and text files in order to present the user with supporting information to the current identification step and the finally derived organism. In *KeyToNature*, a FEDORA (Flexible Extensible Digital Object Repository Architecture) media repository is used to this end as a supplement to multimedia data stored directly in the key database. FEDORA is a conceptual framework that uses a set of abstractions about digital information to provide the basis for software systems that can manage digital information.

In order to make the output of software packages like FRIDA and LINNAEUS compatible and editable in other tools, a certain standard needs to be adopted. *KeyToNature* has decided to adopt the SDD (Structured Descriptive Data) standard proposed by TDWG (Biodiversity Information Standards, formerly Taxonomic Database Working Group). The goal of the SDD standard is to allow capture, transport, caching and archiving of descriptive data in all forms, using a platform- and application-independent, international standard [6].

The keys generated by software packages capable of producing SDD files can then be adapted to produce localized minikeys, for specific applications such as school gardens, parks and reserves, or enhanced with user-generated data by using the Open Key Editor.

Many keys made available by the data providers are 'master keys' including many species. Long keys are complicated and have redundant information when used in an area with fewer species, such as a park or nature reserve, or a school garden. The Open Key Editor allows users to 'crop' a master key and customize it for a given set of species. The 'cropped' key can then be edited for language and illustrations (e.g. to suit a particular user level, or platform such as the mobile phone). With the Open Key Editor the user can browse existing master keys and edit them.

The Open Key Editor has two further important features: 1) it permits to largely solve the problem of translation: once a "large" key has been translated into a given language, it is possible to derive from it a high number of smaller keys adapted to the users' needs without the need of further translations, 2) It permits users to add to the key user-generated content in their own language, thus enhancing considerably the degree of interaction between users and *KeyToNature* identification tools. The Open Key Editor was developed and optimized jointly by the University of Trieste and ETI and is based on open source software. In addition to access on the Internet, output on mobile platforms was included [6].

Keys for mobile devices form further important elements of the system architecture. Many of the keys can be used in the field on PDAs or iPhones, both in stand-alone and in online versions. Tools like the MobilePackager developed by *KeyToNature* partner GIUNTI Labs increase the benefit of the keys by giving users the important possibility of adding user-generated content (in this case geo-referenced pictures) directly in the field, using their mobile devices.

Another useful tool developed in the course of the *KeyToNature* project is IBIS-ID (Interactive Biodiversity Identification Software): it is a "key player software tool" created to help the users in the process of identification of species or other taxa, by using the multi-access keys described in a SDD (Structure of Descriptive Data) file. It is based on the Adobe Flex technology, a well suited

candidate because of its effectiveness for data driven interactive applications and native support for dealing with data organized in XML (eXtensible Markup Language) structured files through the support of the ECMA e4x (ECMAScript for XML) standard [7].

Last but not least it is also possible to comfortably view and edit keys in Wikis, the so called "wiki-keys". *KeyToNature* developed two tools to this end, the jKey player and jKey Editor: The jKey key player (http://www.keytonature.eu/wiki/JKey_Player) is a small javascript, that allows wiki-keys - in addition to their printable overview display - to be also "played" step-by-step, similar to the FRIDA and Open Key functionality. The complementary jKey Editor (http://www.keytonature.eu/wiki/Wiki-based_identification_key_editor) allows form-based editing of the identification keys.

3 A Basic Business Model for Exploitation

A business model is a model of an existing business or a planned future business. A model is always a simplification of the complex reality. It helps to understand the fundamentals of a business or to plan how a future business should look [8].

In the following the key elements of a business model for the possible commercial application of some identification tools and expertise developed within *KeyToNature* is presented:

The value proposition, which describes the key value generated for the main customers and business partners: In our case the key value is supporting customers in solving complex identification problems by utilising expertise and tools of the value chain partners.

Product/market design: The primary end user groups for the *KeyToNature* project are pupils/students and their teachers in formal learning environments. The main products generated in the project are software keys for identifying organisms, which are embedded in learning management systems to provide an adequate context for the respective target group. The formal education market is a difficult one in terms of generating revenues. However, the authors believe that it can be continued to be served for free if maintenance and further development of the software tools can be mainly financed through other commercial activities. Although the identification tools could in principle be applied to any industry with identification problems (e.g. through software keys for identifying bacteria in medicine, or for identifying diseases starting from a series of symptoms) the first step will be to offer the existing tools to a market where there is already relevant domain expertise in the partner network through existing data providers (e.g. plants, certain animals, microfungi). The new business shall thus start off with software keys for desktop and mobile devices in this field as products. The customers could be sought in the B2B area, while individual partners of the network may also address the B2C market directly. Examples of potential customers are nature parks, who want to provide their visitors with specific mobile keys to their flora and fauna. Already in the past it could be proved that this market is worth addressing.

The value creation architecture outlines in which steps the product will be

generated and which partners and competencies are necessary. The following figure outlines the value chain:

Fig. 2 – Value chain of the proposed business model.

When a customer wants support in solving identification problems (for example, a key to the plants in a nature park in Catalonia), a specific software identification tool is provided as a product. It is based upon data provided by one of the data providers who own primary data (i.e. master keys) and secondary data (multimedia files), in our example this might be the Royal Botanical Garden of Madrid with the eflora Iberica database. If the customer cannot provide a list of all species that are in the relevant set, a domain expert (who might be from the Royal Botanical Garden or some other institution), it will be necessary to analyse the situation locally. The master key data together with the list containing the relevant subset of species is then passed on to a key generating software (e.g. the University of Trieste with their FRIDA software or ETI with LINNAEUS, etc.), who produces the specific interactive key. Finally, a developer (e.g. the *KeyToNature* project partner EVOLARIS or Divulgando, a University of Trieste spin-off) creates an application and customized user interface according to the customer's needs and wants (e.g. an iPhone application).

The proposed business model thus draws upon expertise of various partners to solve the problem of the customer. If one of the partners could serve the customer's needs alone, then there is no need for cooperation and the business model doesn't work there. Sometimes this may be the case for data providers, who are at the same time key generator software vendors. But in many cases they may either lack the required master data, resources to build the list containing the subset of relevant species, or the know-how to build the final application with outstanding user experience, e.g. for Apples iPad.

The value architecture could be set up in a way that a newly founded company following the proposed business model could flexibly collaborate with any organisation able to deliver value in one of the four stages through classic buyer-seller-relationships. In the first instance however, the main suppliers of this company could be those *KeyToNature* project partners who decide to join the value creation network.

Revenue model: In order to be able to offer the products and services described above to the market, a company would need to be established, ideally a Ltd. or the like to limit personal liability of the involved individuals in case of potential losses. This company could have the following main objectives: marketing & sales, i.e. finding suitable customers and contracting with them, and sourcing the necessary expertise and tools within the partner network. The company should thus be kept very lean regarding production capacities – here all steps should be outsourced to *KeyToNature* partners and potentially further network members, especially *KeyToNature* associate members. The new company thus would need to manage the partner network to orchestrate the value chain,

negotiate contracts with the customers and ensure proper service through service level agreements. All partners in the network could freely negotiate what they want to get in return for providing their expertise and tools to the company so that the final product could be created. For its services the company should retain the difference between the revenues it could generate from the market and the costs incurred for the services of the business partners.

4 Conclusion

The paper presents tangible outcomes of the *KeyToNature* project and a basic business model for their commercial exploitation. Based on this model a sound business plan with detailed analysis of costs and revenue forecast needs to be developed in order to establish a sustainable business. All *KeyToNature* partners and further interested suppliers of data, expertise, and tools are invited to collaborate with the company, if and when this shall be founded.

Acknowledgement

This work has been supported by the *KeyToNature* project, ECP-2006-EDU-410019, in the frame of the *e*Content*plus* Programme.

References

[1] P. L. Nimis and S. Martellos, "*KeyToNature* a European Project for Teaching Biodiversity". In: A. L. Weitzman and L. Belbin (eds.), *Proceedings of TDWG, Abstracts of the 2007 Annual Conference of the Taxonomic Databases Working Group*, Bratislava, Slovakia, 16-22, p. 67, September 2007.

[2] N. J. Foss (ed.), *Resources, Firms and Strategies: A Reader in the Resource-Based Perspective*. Oxford University Press, Oxford, 1998.

[3] M. E. Porter, *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. The Free Press, New York, 1980.

[4] C. Kittl, *Kundenakzeptanz und Geschäftsrelevanz als Grundlage ökonomisch sinnvoller Geschäftsmodelle für digitale Dienste*, Gabler, 2009.

[5] P. Schalk, P. L. Nimis and W. Addink, "*KeyToNature* Species Identification e-Tools for Education". In: *Biodiversity Information Standards (TDWG) Annual Conference 2008*, http://www.keytonature.eu/w/media/4/4b/KeyToNature_Species_Identification_e-Tools_for_Education_TDWG_2008.pdf, 2008.

[6] P. L. Nimis, N. Dorigo Salamon, C. Kittl, G. Hagedorn and P. Schalk, *D1.6. Annual Report*, 2nd Annual Report to the *e*Content*plus* project ECP-2006-EDU-410019/KeyToNature, 2009.

[7] M. Giurgiu, A. Homodi and G. Hagedorn, "IBIS-ID, an Adobe FLEX based identification tool for SDD-encoded multi-access keys". In: *Biodiversity Information Standards (TDWG) Annual Conference 2009*, 9-13 November 2009, Montpellier, France. http://www.tdwg.org/fileadmin/2009conference/documents/PreProceedings2009.pdf, 2009

[8] P. Stähler, *Geschäftsmodelle in der digitalen Ökonomie: Merkmale, Strategien und Auswirkungen*. Josef Eul, Lohmar, 2002.

Nimis P. L., Vignes Lebbe R. (eds.)
Tools for Identifying Biodiversity: Progress and Problems – pp. 445-450.
ISBN 978-88-8303-295-0. EUT, 2010.

Keys to Nature: A test on the iPhone market

Rodolfo Riccamboni, Alessio Mereu, Chiara Boscarol

Abstract — Several keys running on mobile devices were developed by the *KeyToNature* project. Most of them are freely downoadable online. The rapid spread of smartphones has opened up new opportunities in the production and distribution of multimedia applications for the educational sector, including interactive keys to identify organisms. This market is new, still partly unexplored and changing fast. The Department of Life Sciences of the University of Trieste and Divulgando Srl, have tested its potential by uploading in the iTunes Store different types of keys for the iPhone, some of them for free, others for sale. This paper introduces the issue of global market applications, summarizes the experience gained in our case-studies, and suggests ways to make these applications economically viable.

Index Terms — smartphones, mobile store, iPhone applications, education, global market, Apple, Android.

1 Introduction

One of the most successful products developed by *KeyToNature* are identification keys to plants, animals and fungi running on mobile devices (PDAs and smartphones). Since they can be used in the field, they proved to be useful in schools, also attracting the attention of many associate members of *KeyToNature* such as Natural Parks and Botanic Gardens, as a means to advertise their biodiversity heritage. Presently, several hundreds of applications for PDAs, most of which specifically created for a single school, are freely downloadable online from the Italian portal of *KeyToNature* (www.dryades.eu).

The rapid spread of smartphones has opened up new opportunities in the production and distribution of multimedia applications for the educational sector, including interactive keys to identify organisms. This market is new, still

R. Riccamboni is with the Department of Life Sciences, University of Trieste, I-34127, Italy. E-mail: rricamboni@units.it.
A. Mereu is President of Divulgando S.r.l., Corso Italia, 31, I-34122, Trieste, Italy. E-mail: mereu@divulgando.eu.
C.Boscarol is Content Manager of Divulgando S.r.l., Corso Italia, 31, I-34122, Trieste, Italy. E-mail: boscarol@divulgando.eu.

partly unexplored and changing fast. The Department of Life Sciences of the University of Trieste and Divulgando Srl, have tested its potential by uploading in the iTunes Store different types of keys for the iPhone, some of them for free, others for sale.

This paper introduces the issue of global market applications for mobile devices, summarizes the experience gained through four case-studies, and suggests ways to make these applications economically viable.

2 The market of mobile applications

In the past 12 months, the global market for applications has greatly increased its sales, mainly due to Apple Inc. and its iPhone App Store. At the end of April 2009, downloads from the App Store exceeded 1 billion, a year later they were over 4 billions [1]. This market leadership is due to several factors:

1. developers create applications for a single operating system running on all devices in the iOS market (something that happens on Nokia Symbian as well),
2. procedures for purchasing an application are very simple,
3. the hardware is of excellent quality, the software is very user-friendly,
4. the process is reversed: people buy the iPhone because there are many iPhone applications,
5. the satisfaction of customers buying the iPhone is 73% compared with 39% of HTC (High Tech Computer Corporation) [2].

Juniper Research, an International Company of statistics and market forecasts, states that the App Store market - worth nearly $ 10 billion in 2009 - will be worth over $ 32 billion in 2015, with an increase of over $ 22 billion dollars in 6 years [1]. The development of mobile applications is not an ephemeral fashion, but a strong reason for business and visibility. A challenge against Apple's App Store is now posed by all the major mobile phone companies, such as Samsung Rim, Microsoft, Nokia, and Intel. Currently, developers prefer the Apple App Store of Apple, but Google Android is rapidly increasing in terms of both applications and devices sold.

3. Aim, data and methods

We tried to assess the potential of the market for iPhone applications by developing and making available on the App Store of iTunes several identification tools to vascular plants developed by *KeyToNature*, differing in:

- geographic coverage (nationwide to very local),
- number of species involved (from c. 100 to more than 2000),
- dimension of market (which mainly depends on language: Estonia vs. Italy-The Netherlands),
- price (for sale vs. for free).

All applications were developed by the University of Trieste and Divulgando Srl, except one, developed by ETI Bioinformatics. The results are based on the absolute number of downloads for each type of application sold, and on their relative rank in the educational sector of the respective national markets

(which varies daily depending on the number of downloads). This information is provided to developers by Apple Inc. every day. The ranking of applications is related to both the category where they are included (e.g. entertainment, travel, games, education) and to the total number of applications sold in each category. Each country has its own history each store has its own ranking.

Developers of applications in the Apple App Store can change the pricing of an application almost in real time: we took advantage of this opportunity by changing the price of some applications for short periods of times, to test whether the price had an influence on the number of downloads.

4 Case-study applications

4.1 Woody plants of Estonia (Eesti eFloora I. Puud ja põõsad)

This key to over 140 trees and shrubs growing in Estonia [3] was our first testbed. The key is dichotomous, richly ilustrated. It includes taxon pages with notes (in Estonian), distribution maps and pictures of every species. It can also interacts with users, who can add user-generated content in "their" own key in the form of textual field-notes and pictures, and permits users to share their input in the Web community (Facebook).

With this key we wanted to test the iPhone market in the worst conditions:

- in a very small market (the key is written in Estonian),
- in an unfavourable period (January 2010, when most of Estonia was covered by snow),
- charging a fixed cost of € 1.59.

The result was surprising: in a couple of days this became the best-seller in the educational sector of the App Store in Estonia. Still now, at a distance of 8 months, it still keeps a leading position among the 10 bestellers in this sector. The total number of downloads (c. 370) is small, but it ranks high in the educational sector of the Estonian market. A peak with over 120 downloads (almost 1/3 of the total) was recorded on May 13th, 2010, on the occasion of a national TV service, which shows how important advertising is for the iPhone market.

Lesson for us: the market for identification keys on mobiles is interesting.

4.2 The flora of the Netherlands

This application, produced by another partner of *KeyToNature* - ETI Bioinformatics - is an interactive flora with a nationwide coverage, including more than 1.800 species, which manages a large amount of data (375 Mb). It was uploaded in the Apple App Store in June 2010, at the rather high price of 9.99 €. Apple Inc. has set a threshold of 20 Mb for direct download on the iPhone via UMTS/GPRS. Applications that exceed this limit must be installed by connecting to a computer, and in general lightweight applications (less than 20 Mb) tend to be downloaded more than those exceeding the threshold of 20 Mb. In spite of this and of the rather high price, the market responded well: from June

15th to August 15th the key had ca. 700 downloads.

Lesson for us: the market for identification keys is not only interesting, but also economically rentable.

4.3 A "very local" key for free in the Italian market (100 plants in the Botanical Garden of Catania - Sicily)

At this point we wanted to test the potential of another large market, that of Italy. The first experiment was based on a very local-special key, that to 100 woody plants occurring in the Botanical Garden of Catania (Sicily). Prepared in collaboration with colleagues from the University of Catania, this key was originally meant to be used only for educational activities organised by the Garden. We thought that it would be of little interest on the Italian iPhone market, but the results were surprising: the application was made available for free on the App Store in May 2010: in the first 20 days it had over 1000 downloads, and still now (August) is having 20-40 downloads a day. On August 2, 2010 the total downloads were more than 4000.

Lesson for us: the market for identification keys in Italy is very interesting.

4.4 Other "Local" keys for sale in the Italian market (Flora of the Trieste Karst area, NE Italy)

The previous lessons told us that the Italian market is wide and economically interesting. Thus, we tried to "sell" on the Apple App Store some local keys to plants, those occurring in the Karst area near Trieste, a small enclave at the eastern border of Italy which however is poorly known by Italians.

The test was based on 2 keys:

- a "large" key to c. 1000 species occurring in the Val Rosandra Natural Reserve near Trieste, sold at 2.39 €
- two smaller keys (200-250 species) to the plants of special habitats in the Karst area (dry grasslands, woody habitats), sold at 1.59 Euros.

These keys were uploaded on iTunes at the end of June 2010.

The general results were disappointing: the smaller keys had a constant average of c. 3 daily downloads, while the downloads of the larger key were even smaller (average 1.5 a day). At this point we have lowered the price of all keys to 0.79 Euros, without any relevant change in the number of downloads. Finally, we made the applications downoadable for free. The experiment lasted 3 days only (August, 18-20). The result was surprising: in 3 days the number of daily downloads rose dramatically, with an average of 60 downloads a day per key and an increasing trend. There was a significant difference in the number of daily downloads between the two very similar smaller keys (dry grasslands and woody habitats). The latter had many more downloads a day.This difference may be due to the titles of the 2 keys: the former was entitled "La landa carsica", where "landa" is a term for dry grasslands which is used only locally in the Karst area, while the second was entitled "Piante di sottobosco" (plants of the understorey), where the term "sottobosco" is well known nationwide. This suggests to pay attention to title and keywords when uploading a key on the

Apple App Store.

Lesson for us: "local" keys have no market if sold for money, but may have a large impact if made available for free.

5 Conclusions

The results of our tests are summarised in Tab. 1. The most relevant conclusions from our tests are:

KEY	GEOGRAPHIC COVERAGE	NUMBER OF SPECIES	MARKET	PRICING	Nr. of DOWNLOADS
Estonia	nationwide	medium	small	for sale	**very high**
Flora of the Netherlands	nationwide	very high	wide	for sale	**very high**
Catania Botanical Garden	very local	small	wide	free	**very high**
Karst flora	local	small	wide	free	**very high**
Flora of the Val Rosandra Reserve	very local	high	wide	free	**high**
Karst flora	local	small	wide	for sale	**low**
Flora of the Val Rosandra Reserve	very local	high	wide	for sale	**very low**

Tab. 1 – Synthesis of the results.

1. The interest for identification tools in the iPhone market is potentially high.
2. The geographic coverage of a key has a great impact on the market: nationwide keys can be profitably sold, while local keys seem to be a poor source of revenues.
3. However, when local keys are made downloadable for free, their impact increases dramatically, and they are downloaded also by persons outside of the area of interest of the key.

On the light of these considerations, we have changed our market strategy as follows:

1. We have started the production of keys with a nationwide coverage, which will be placed on the market for sale.
2. We are proposing to the many Parks and Natural Reserves for which we have already created a key a further service, that of developing an application for the iPhone, which will be made available for free as a powerful means to advertise their biodiversity heritage. Several Parks ready showed a keen interest, and are redy to pay for such a service.

Acknowledgement

This work has been supported by the *KeyToNature* project, ECP-2006-EDU-410019, in the frame of the *e*Content*plus* Programme.

References

[1] W. Holden, *A World of Apps*. Whitepaper from Mobile App Stores, Business Model, Strategies & Market Segmentation 2010-2015, Juniper Research Ltd, June 2010.

[2] J. Crumrine and P. Carton. Explosive Changes in Consumer Demand Shake Up Smart Phone Industry. ChangeWave Research Ltd. http://www.changewaveresearch.com/articles/2010/07/smart_phones_20100714.html, July 14, 2010.

[3] A. Saag, T. Randlane and M. Leht, "Key to plants and lichens on smartphones: Estonian examples". In: P. L. Nimis and R. Vignes Lebbe (eds.), *Tools for Identifying Biodiversity: Progress and Problems*, pp. 195-199, 2010.

Author Index

A

Addamiano S., 157
Addink W., 55
Aedo C., 99, 177
Agoo E. M. G., 341
Anastasi A., 183
Aplikioti M., 231
Arena S., 327
Argyrou M., 231
Assunçao, O., 419
Atanacio R., 31

B

Bailly N., 31
Balenghien T., 201
Balke M., 347
Barberá P., 99
Barbouche N., 343
Barthélémy D., 221, 237
Basheer V. S., 345, 353
Baylac M., 343
Berendsohn W. G., 1, 7
Bertolani R., 333
Bineesh K. K., 345
Bisby F. A., 37
Blanco G., 281
Boar F., 367
Bonnet P., 221, 237
Borg J. J., 25
Borrell Y. J., 281
Boscarol C., 445
Boujemaa N., 237
Brandner R., 13
Brewington S. D., 171
Buono S., 249

C

Cabezas F., 99
Campo D., 315
Candolfi E., 201
Cano Calonge C., 163
Caron D., 217
Casiraghi M., 269, 289
Causse F., 113, 383
Cazet J.-P., 383
Cesari M., 333
Cêtre-Sossah C., 201
Chabalier J., 419
Chanthavongsa K., 221
Chavernac D., 201
Chiatante D., 327
Chouchene M., 343
Combourieu Nebout N., 383
Conruyt N., 65, 71, 217
Coullet O., 419
Crego-Prieto V., 315
Crucianu M., 237
Cubelio S. S., 345
Cuming D., 423

D

De Giovanni M., 25

De la Estrella M., 99

De Mattia F., 269

Delécolle J. -C., 201

Diep M. H, 207

Ditzler S., 349

Dorigo Salamon N., 437

Drakos A., 231

Dupont J., 383

E

Esparza I., 323

F

Faure G., 217

Fero M., 99

Ferrer M., 373

Ferri E., 269

Filipello Marchisio V., 183

Fortini P., 257

Fradin H., 7

Froese R., 31

Froufe E., 301

Furlan M., 307

G

Galán P., 323

Galimberti A., 269, 289

Gallagher M. S., 417

Gamarra R., 323

García E., 373

Garcia-Vazquez E., 315

Garnery L., 343

Garros C., 201

Geiger M., 351

Gérard D., 107

Geynet Y., 217

Giurgiu M., 13, 19, 83, 133, 137

Gopalakrishnan A., 345, 353

Graciete L., 219

Gransinigh E., 249

Grard P., 221

Graziosi G., 307

Grosser D., 65, 71, 217

Grube M., 295

Guarino R., 157, 405

Gudmundsson G., 171

Güntsch A., 7

H

Hagedorn G., 13, 19, 59, 77, 83, 89

Haszprunar G., 347

Häuser C. L., 421

Hausmann A., 347

Hebert P., 347

Hendrich L., 347

Hetzner S., 77, 137

Hoffmann A., 421

Hoffmann J., 43

Hoffmann N., 7

Hominick B., 429

Homodi A., 13, 19, 83, 133

Hopkins D., 395, 401

I

Ialicicco M., 327

J

James K., 395

Jong Y. de, 49

K

Kastrantas K., 411

Kerekes A., 367

Kirchhoff A., 7

Kirmitzoglou I., 231

Kittl C., 13, 361, 437

Kizhakudan J., 353

Kizhakudan S. J., 353

Knebelsberger T., 349

Kodele Krašna I., 379

Kohlbecker A., 7

Kroupa A., 421

Kuntzelmann E., 7

L

La Rosa M., 157

Laakmann S., 349

Labra M., 269

Lakra W. S., 345, 353

Lamxay V., 221

Lanorsavanh S., 221

Laroche P., 223

Larsen K., 301

Leht M., 189, 195

Lopes M., 219

Loy A., 243, 257, 263

M

MacLeod N., 225

Magrini S., 249, 251

Maiocco Ô., 7

Manouselis N., 411

Martellos S., 13, 59, 77, 115, 127, 133, 151, 437

Martin P. A., 65

Martinoli A., 289

Mathieu B., 201

Mathieu D., 121

Maxl E., 361

McGovern T., 171

Medina L., 177

Menegoni P., 405

Mereu A., 445

Mihnev P., 13, 355

Mohrbeck I., 349

Monje J. C., 421

Montaño A. J., 163

Morris R. A., 13

Mucedda M., 289

Muggia L., 295

Mulder P., 209

Müller A., 7

Mulrenin A., 25

N

Nascimbene J., 151

Neumann F., 137

Nguyen B. L., 207
Nguyen H. P., 207
Nicolosi P., 263
Nikolaou N., 231
Nimis P. L., 13, 19, 77, 127, 133, 151

O

Onofri S., 249
Ortúñez E., 323
Ouertani W., 237

P

Palavitsinis N., 411
Pallavicini A., 307
Papamarkos N., 231
Perez J., 315
Pernot T., 121
Peters P., 419
Petersen A., 25, 171
Pieterse S., 25
Pignatti S., 157, 405
Pinho R.M., 219
Plank A., 13, 77
Press B., 77, 389
Promponas V. J., 231

R

Raj K., 345
Rakhee C., 353
Ralambondrainy H., 71
Rambold G., 59
Randlane T., 189, 195
Raupach M. J., 349
Raycheva N., 355
Rebecchi L., 333
Rekha J. N., 353
Rempicci M., 249
Reyes R. Jr., 31
Riccamboni R., 445
Richardson A., 401
Roberts D., 53
Rocco M., 327
Romano F., 281
Roskov Y. R., 37
Roujinov M., 13
Rovellotti O., 419
Russo D., 289

S

Saag A., 189, 195
Sahl A., 419
Sajeela K. A., 353
Sampaziotis P., 231
Sánchez Laulhé F., 163
Sánchez Prado J. A., 281
Sanz E., 323
Sbordoni V., 275
Scaloni A., 327
Schalk P., 13, 55, 127, 429, 437
Schmidt G., 13, 137
Schmidt S., 347
Scholz H., 43
Schuiteman A., 221

Scippa G. S., 327
Scoppola A., 249, 251
Seijts D., 127
Silva M. H., 219
Silveira P., 219
Slice D. E., 243
Smith V., 53
Steinmann R., 25
Stoitsis J., 411
Strasser A., 25
Straube N., 351
Svengsuksa B., 221

T

Talbi K., 419
Targetti S., 423
Tarkus A., 361
Teage I., 25
Tewari S., 345
Tornincasa P., 307
Trayler A., 25
Triebel D., 13
Trilar T., 95
Trupiano D., 327
Tsilibaris X., 411

U

Uiterwijk M., 213
Ung V., 113, 201

V

van Raamsdonk L. W. D., 145, 213
van Spronsen E., 13, 55, 127, 133
Varese G. C., 183
Vázquez E., 281
Veja C., 13, 19
Velayos M., 99
Venin M., 7
Viaggi D., 423
Vignes Lebbe R., 7, 107, 113, 201, 207, 383
Viscosi V., 257, 327
von Mering S., 77
Voyron S., 183

W

Weber G., 13, 77, 89

Z

Zammit N., 25
Zelazny B., 13

SIXTH FRAMEWORK
PROGRAMME

The organising Projects are co-funded by the European Union:
*e*Content*plus* Programme and Sixth Framework Programme (FP6)

2010 International Year of Biodiversity

In cooperation with UNESCO Venice Office for Science and Culture in Europe

MUSEUM NATIONAL D'HISTOIRE NATURELLE

In cooperation with Muséum national d'Histoire naturelle, Paris

ISBN 978-88-8303-295-0

9 788883 032950